Using a genetic algorithm to estimate the details of earthquake slip distributions from point surface displacements

AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Solid Earth

RESEARCH ARTICLE
10.1002/2015JB012181

Key Points:
- Details the implementation of a genetic algorithm for constraining earthquake slip distributions
- Uses coral displacement data to estimate the slip distributions for recent and paleoearthquakes
- Solutions identified by the technique may be useful in slip deficit and stress evolution modeling

Supporting Information:
- Supporting Information S1

Correspondence to:
A. Lindsay,
Lindsay-a5@email.ulater.ac.uk

Citation:
Lindsay, A., J. McCloskey, and M. Nic Bhloscaidh (2016), Using a genetic algorithm to estimate the details of earthquake slip distributions from point surface displacements, *J. Geophys. Res. Solid Earth*, *121*, 1796–1820, doi:10.1002/2015JB012181.

Received 8 MAY 2015
Accepted 26 JAN 2016
Accepted article online 29 JAN 2016
Published online 12 MAR 2016

Using a genetic algorithm to estimate the details of earthquake slip distributions from point surface displacements

A. Lindsay[1], J. McCloskey[1], and M. Nic Bhloscaidh[1]

[1]Geophysics Department, School of Environmental Sciences, University of Ulster, Coleraine, UK

Abstract Examining fault activity over several earthquake cycles is necessary for long-term modeling of the fault strain budget and stress state. While this requires knowledge of coseismic slip distributions for successive earthquakes along the fault, these exist only for the most recent events. However, overlying the Sunda Trench, sparsely distributed coral microatolls are sensitive to tectonically induced changes in relative sea levels and provide a century-spanning paleogeodetic and paleoseismic record. Here we present a new technique called the Genetic Algorithm Slip Estimator to constrain slip distributions from observed surface deformations of corals. We identify a suite of models consistent with the observations, and from them we compute an ensemble estimate of the causative slip. We systematically test our technique using synthetic data. Applying the technique to observed coral displacements for the 2005 Nias-Simeulue earthquake and 2007 Mentawai sequence, we reproduce key features of slip present in previously published inversions such as the magnitude and location of slip asperities. From the displacement data available for the 1797 and 1833 Mentawai earthquakes, we present slip estimates reproducing observed displacements. The areas of highest modeled slip in the paleoearthquake are nonoverlapping, and our solutions appear to tile the plate interface, complementing one another. This observation is supported by the complex rupture pattern of the 2007 Mentawai sequence, underlining the need to examine earthquake occurrence through long-term strain budget and stress modeling. Although developed to estimate earthquake slip, the technique is readily adaptable for a wider range of applications.

1. Introduction

To properly assess the seismic hazard in prone areas, we need to understand the processes influencing the location and magnitude of future earthquakes [*Sieh et al.*, 2008; *Nalbant et al.*, 2013]. However, the factors controlling earthquake recurrence are poorly understood. Examining the accumulation and release of interseismic stresses over a number of seismic cycles may provide insight into these mechanisms.

While we can determine secular stress loading rates from the relative motions of tectonic plates [*Bock et al.*, 2003], it is the physical properties of the fault that control where the stress is stored. Features of the plate interface, which we refer to as physical asperities, such as topographic roughness, fracture zones, bathymetric features (i.e., sea mountains and plateaus), overlying landforms, and other frictional properties interact to prevent continuous aseismic plate movement [*Pacheco et al.*, 1993; *Chlieh et al.*, 2008; *Kaneko et al.*, 2010]. Over time, where physical asperities prevent free movement along sections of the fault, relative to the ratio of plate motion, we observe the accumulation of a slip deficit. The degree of coupling describes the extent to which physical asperities limit free movement of the plates and is defined by *Chlieh et al.* [2008] as the ratio between the slip rate deficit and the long-term slip rate. We observe aseismic creep along weakly coupled sections of the plate interface. Conversely, where there is strong coupling, the accrual of interseismic strain results in an elastic deformation of the interface [*Chlieh et al.*, 2008; *Konca et al.*, 2008; *Moreno et al.*, 2010; *Kaneko et al.*, 2010].

The size of the slip deficit limits the amount of potential slip that can occur in future earthquakes [*Chlieh et al.*, 2008]. Coseismic relaxation of the plate interface results in the heterogeneous release of accumulated strain as a patchwork of relatively high and low slip regions, referred to as "slip asperities" and "antiasperities," respectively [*Chlieh et al.*, 2008]. The correlation between strongly coupled sections of the plate interface and slip asperities has been observed in previous studies [*Chlieh et al.*, 2008; *Konca et al.*, 2008; *Moreno et al.*, 2010; *Lorito et al.*, 2011]. For example, along the Nias-Simeulue section of the Sunda Trench, areas of high coupling correspond well with the location of slip asperities [*Konca et al.*, 2008]. However, farther south, along the Mentawai section of the trench, in a sequence of earthquakes that occurred in 2007 (Figure 1a) the

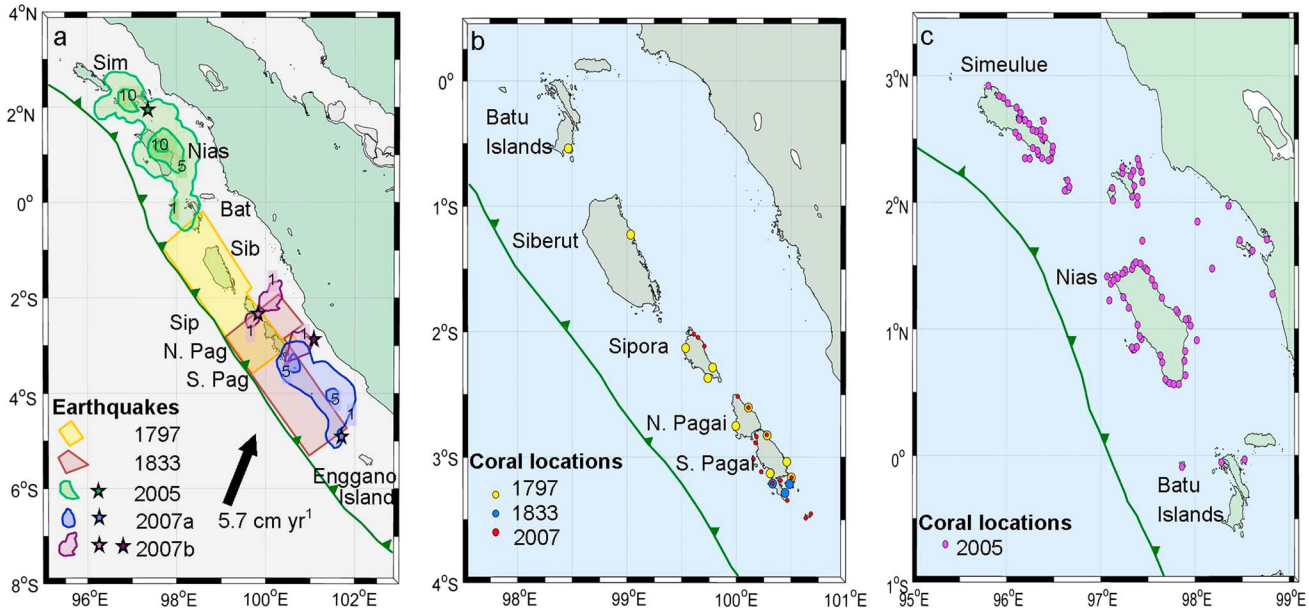

Figure 1. (a) The Outer Arc Islands of the Sunda Trench with outlines of a number of recent and paleoearthquakes. Source parameters of the 1797 (yellow) and 1833 (purple) earthquakes are taken from *Natawidjaja et al.* [2006]. The 1 m, 5 m, and 10 m slip contours of the 2005 M_w 8.6 Nias-Simeulue earthquake are shown in green with a green star marking the event epicenter [*Konca et al.*, 2007]. The 1 m and 5 m slip contours for the 2007 M_w 8.4 primary rupture are shown in blue, and the 1 m contours for the later M_w 7.9 sequence are shown in purple with correspondingly colored stars making the event epicenters [*Konca et al.*, 2008]. (b) The Mentawai section of the Outer Island Arc showing the location of the coral microatolls recording uplift during the 1797 and 1833 earthquakes and during the 2007 sequence of events. (c) The Nias-Simeulue section of the Outer Island Arc showing coral microatolls recording uplift during the 2005 earthquake.

coupling distribution provides an inadequate guide to the location of the slip asperities [*Konca et al.*, 2008; *Nalbant et al.*, 2013].

We may gain a better understanding of the relationship between physical and slip asperities by examining the release of strain over hundreds of years. For recent earthquakes, the development of GPS and interferometric synthetic aperture radar (InSAR) technologies, first used to study the 1989 Loma Prieta earthquake [*Lisowski et al.*, 1990] and the 1992 Landers earthquake [*Massonnet et al.*, 1993; *Zebker et al.*, 1994], allows detailed inversions of earthquake slip distributions to be produced. However, to examine the details of slip during earlier earthquakes, we need to interrogate other sources of data [*Nalbant et al.*, 2013; *Nic Bhloscaidh et al.*, 2015]. One such source is the coral microatolls growing along the intertidal reefs that fringe the Outer Arc Islands of the Sunda Trench (Figures 1b and 1c). Along the Sunda Trench, the variations in relative sea levels around the Outer Arc Islands are driven by tectonic flexing of the accretionary wedge. The response of coral microatolls to these changes in sea level is recorded in their annual growth rings [*Taylor et al.*, 1987; *Zachariasen et al.*, 1999]. *Taylor et al.* [1987] first demonstrated the use of coral stratigraphy to study tectonic deformation in the Vanuatu Islands. Subsequently, the technique has been applied to and developed for the coral microatolls overlying the Sunda Trench where both living and fossil coral heads store a record of tectonic activity associated with several large earthquakes spanning hundreds of years [*Zachariasen et al.*, 1999; *Briggs et al.*, 2006; *Natawidjaja et al.*, 2006, 2007; *Konca et al.*, 2007, 2008; *Sieh et al.*, 2008; *Meltzner et al.*, 2010; *Philibosian et al.*, 2014]. The coral displacement data have been used to constrain slip distributions for the 2005 and 2007 earthquakes [*Briggs et al.*, 2006; *Konca et al.*, 2007, 2008]. In conjunction with historical reports of shaking and tsunami damage [*Newcomb and McCann*, 1987], the coral record has also helped constrain the magnitude, source parameters, and maximum slip values of two large preinstrumental earthquakes in 1797 and 1833 [*Natawidjaja et al.*, 2006].

The corals of the Outer Arc Islands act as long-term geodetic recorders, and the proximity of the islands both to strongly coupled sections of the plate interface and to the slip asperities of recent earthquakes make them an ideal site to study rupture patterns of large megathrust events. The islands are found between 80 km and 130 km from the trench, approximately 15 km to 25 km above the plate interface (Figure 1a) [*Kopp et al.*, 2001; *Singh et al.*, 2011; *Collings et al.*, 2012]. For earthquakes in 2005 and 2007, researchers were able to combine a number of data sets to produce well-resolved slip inversions, which we refer to here as the Kon05 and the

Kon07 models, respectively. In addition to coral displacement data, they used information from previously installed field and campaign GPS stations, teleseismic data, and, for the 2007 event only, InSAR satellite measurements [*Briggs et al*., 2006; *Konca et al*., 2007, 2008]. The earthquake SouRCe MODel (SCRMOD) database [*Mai and Thingbaijam*, 2014] archives the finite source models for a number of earthquakes and, thus, allows us to compare the Kon05 model to preliminary teleseismic models of the 2005 earthquake [*Shao and Ji*, 2005; *Ji*, 2005]. While there is some variation in the location of their slip asperities, both the Kon05 and the teleseismic models of the 2005 event have similarly ranged slip values. However, the Kon05 joint inversion presents a more compact slip distribution and smaller M_w of 8.5 compared to the M_w of 8.7 produced from the teleseismic data alone. We can also compare the Kon07 model to both preliminary teleseismic models of the 2007 earthquake [*Konca*, 2007; *Ji*, 2007] and a model of the event produced from teleseismic and tide gauge data [*Gusman et al*., 2010]. The teleseismic models of the 2007 event have peak slip values of $<$3.5 m, less than half that of the Kon07 inversion. The distribution of slip and range of slip values in the Kon07 model and the *Gusman et al*. [2010] model are broadly similar, but the more spatially compact slip of the former model has a slightly lower M_w of 8.4 compared to M_w of 8.5 of the latter one. *Natawidjaja et al*. [2006] demonstrated the possibility of using coral data to examine the slip distributions of preinstrumental earthquakes. They produced rectilinear source models of the 1797 and 1833 Mentawai earthquakes where each model was subdivided into four and three blocks of uniform slip, respectively. However, subsequent attempts to use these models to examine the evolving stress state of the Sunda Trench found them to be insufficiently detailed [*Nalbant et al*., 2013]. In response to calls for better resolved models of these events, we have developed a novel Genetic Algorithm Sip Estimation (GASE) technique to constrain the details of earthquake slip distributions from sparse, surface displacements.

This paper describes the design, implementation, and testing of the GASE technique. In section 2, we detail the steps necessary to set up a fault model and define a set of model slip distributions, before describing the implementation of the GASE technique. In section 3, we use synthetic data to test the sensitivity of solutions to initial model constraints. In section 4, we examine estimates of the slip distributions for the 2005 Nias-Simeulue earthquake and 2007 Mentawai sequence from the published displacement data [*Konca et al*., 2007, 2008] and compare our solutions to the published inversions. Finally, we produce the slip distributions for the 1797 and 1833 paleoearthquakes using displacement data published by *Natawidjaja et al*. [2006] and assess the reliability of our solutions. Extensive systematic tests of the implementation and optimization of the techniques have been included in the supporting information for the interested reader.

2. A New Genetic Algorithm Slip Estimator Technique

2.1. Implementation of the GASE Technique

To constrain details of the slip responsible for a set of observed coral displacements, we first construct a fault model discretized into cells of the desired resolution that allows the surface effects of slip to be calculated. We then produce a set of model slip distributions and evaluate each of them on their ability to satisfy the data. Next, we use a genetic algorithm to recombine the information in these models and converge on a set of solutions that better reproduces the coral displacements. The sparse distribution of observations presents an underdetermined problem and infinite number of solutions. We therefore repeat this process for a number of different sets of models; each time storing the model most closely converged on the observations. Finally, from the stored models we compute an ensemble estimate of the slip distribution most likely responsible for the observed coral displacements.

2.1.1. Construction of a Fault Model

Our planar fault model is based on the geometry of the Sunda trench. The origin point of the fault model is at 7°S and 102°E, approximately 160 km south west of Enggano Island (Figure 1). It extends 1280 km northward along the strike of 325° and 240 km downdip to accommodate the full size of the coupling distribution for the Sumatra megathrust proposed by *Chlieh et al*. [2008]. We choose a dip angle (δ) of 15° to allow comparison of the GASE solutions to the combined inversion of geodetic and teleseismic data for the 2007 Mentawai sequence [*Konca et al*., 2008] and an earlier attempt at modeling the 1797 and 1833 paleoearthquakes [*Natawidjaja et al*., 2006]. The fault model is discretized into a total of 768 cells, 64 along strike and 12 downdip, each of 20 km length (ε). A schematic (Figure 2h) shows the fault model with the overlying surface accommodating observation points analogous to the 1797 coral data set and a vertical arrow marking the coral displacement at the location *k*.

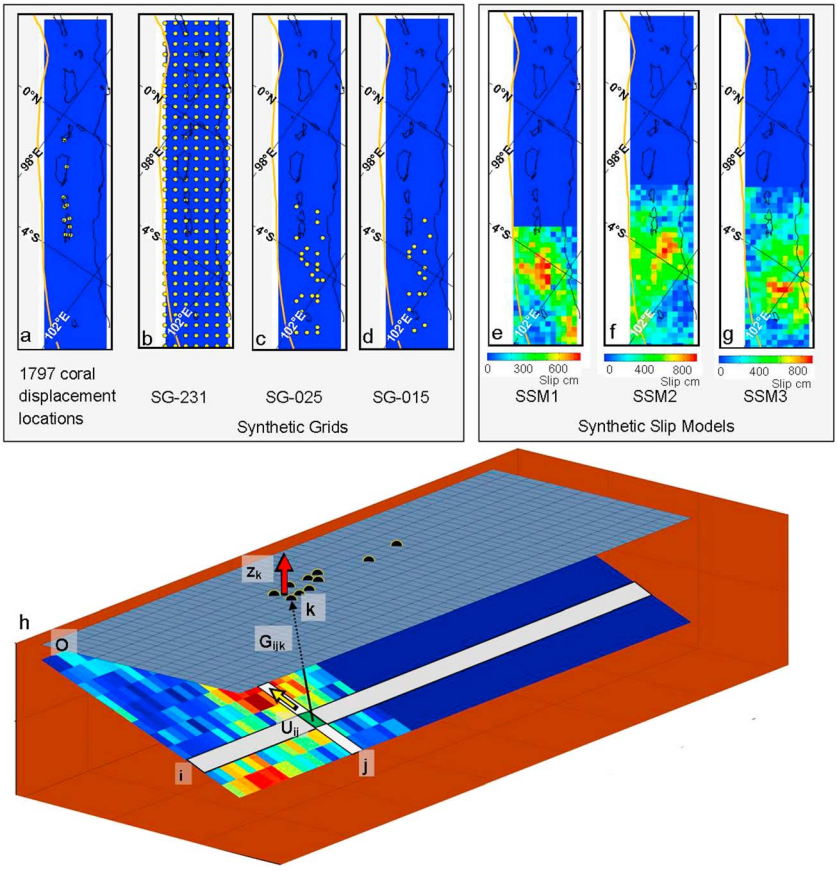

Figure 2. A schematic of the fault model showing the trench lineament and the position of islands together with the distribution of surface observation points (Figures 2a–2d) and overlaying synthetic slip models (Figures 2e–2h): (a) the location of coral microatolls recording uplift associated with the 1797 earthquake; (b) 231 observation points uniformly spaced (40 km) over the whole surface of the fault model (SG-231); (c) 25 observation points nonuniformly spread over the lower 640 km section of the fault plane (SG-025); (d) a subset of 15 observation points taken from SG-025 (SG-015); (e) the synthetic slip model of M_w 8.71 (SSM1); (f) the synthetic slip model of M_w 8.81 (SSM2); (g) the synthetic slip model of M_w 8.78 (SSM3); and (h) an oblique 3-D view of our discretized fault model showing heterogeneous slip to be modeled as vertical surface displacements. A Green's function is calculated by mapping the slip *U*, in each cell *ij*, as a vertical displacement at a surface location *k*. The total vertical displacement at the surface location is found by summing the individual contributions from each cell. The origin point O of the fault model is 7°S, 102°E, with a strike of 325° and dip of 15°. The fault model cell size is 20 km by 20 km, extending 64 cells along strike and 12 cells downdip. Observation points analogous to the 1797 coral locations are shown in black on the surface overlying the fault interface.

We use Gomberg's 3d-def software [*Gomberg and Ellis*, 1993], which incorporates *Okada*'s [1992], to define a Green's function *G* describing the vertical displacement *z* at *k* produced by reverse slip *U* in fault model cell *ij*. While the convergence of the Australian, Indian, and Sunda plates introduces a slight oblique component to the strain accumulation [*Chlieh et al*., 2008], it is not possible to estimate the strike-parallel components of the slip from the vertical displacement data recorded by the corals. As such, our models can only account for reverse slip on the fault; however, it has been shown that the oblique component of coseismic slip for the 2005 Nias-Simeulue and the 2007 Mentawai sequence is relatively small [*Konca et al*., 2007, 2008]. In effect, the total displacement Z_k for each cell of the fault model is defined as

$$Z_k = \sum_{ij} G_{ijk} . U_{ij} \tag{1}$$

For each observation point *k*, a displacement matrix is created with the same dimensions as the fault model. The displacement matrix records the uplift at location *k* produced by 1 cm slip on the equivalent cell of the fault model. The total uplift at *k* is calculated by summing the product of the slip value in each cell of the fault model and uplift value in the corresponding cell of the displacement matrix. Given a set of observation

locations, we compute a displacement matrix for each, allowing the rapid calculation of the vertical deformation resulting from any heterogeneous model slip distribution. Note that throughout the paper, we consistently use the term "slip" to describe motion along the fault interface and the term "displacement" to refer to the vertical surface displacement.

Next, we evaluate the goodness of fit f_k for each solution from the mismatch between each observed displacement Z_{ok} and the corresponding modeled displacement Z_{mk} weighted by the observational uncertainty at each location γ_k:

$$f_k = e^{-\frac{1}{2}\left[\frac{(Z_{ok}-Z_{mk})^2}{\gamma_k^2}\right]} \tag{2}$$

For each observation, f_k has a value between 1, indicating a perfect fit to the data, and 0, indicating a poor fit. The total fitness F for any model solution is obtained from the product of f_k, $k = 1 \ldots n$ where n is the total number of observation points:

$$F = \prod_k^n f_k \tag{3}$$

In order to compare the fitness values of solutions based on different numbers of observation points, we calculate the geometric mean taking the nth root of the product of n observation points:

$$F' = F^{\frac{1}{n}} \tag{4}$$

This likelihood-based method severely penalizes models that are a poor fit to the data, allowing us to extract the maximum amount of information from sparse sets of observations.

2.1.2. Heterogeneous Slip Model Constraints

While we can increase the detail at which we model the earthquake slip distributions by reducing the cell size in the fault model, the improved model resolution comes with increased computational overheads. The number of dimensions required to describe the solution rises with the number of fault model cells and requires us to consider ways of constraining the range of possible models without excluding any viable solutions. The model space required to accommodate the full range of continuous slip values in the 768-cell fault model is infinite, yet the majority of produced models constitute unphysical simulations of earthquake slip, i.e., models where the large difference between slip values in neighboring cells will result in the fracturing of the rock rather than producing an elastic deformation. We can produce realistic, heterogeneous earthquake slip distributions by sampling slip values from a fractal distribution. This allows us to reproduce the power law decay exhibited by the power spectrum of other finite-fault source inversions of earthquake slip [*Mai and Beroza*, 2002] and to replicate the self-affine, scale-invariant properties observed in ruptures [*Turcotte*, 1997]. We can vary the slip heterogeneity, and so the proximity of slip asperities to antiasperities, by varying the fractal dimension of the fractal distribution. For example, where the fractal dimension equals 2.1, it allows 10 m of slip to fall to 0 m over 100 km [*McCloskey et al*., 2008]. By decreasing the fractal dimension to 1.5, the slip variability is significantly reduced. It results in smoother appearing earthquake models, with slip falling, on average, from 2.5 m to 0 m over 100 km. On the other hand, increasing the fractal dimension raises the variability of slip values in adjacent fault model cells, so that a fractal dimension of 2.8 allows slip of 20 m to fall to 0 m over 100 km. Therefore, by accepting the slip distributions with the fractal dimensions in the range 2.0 to 2.5, we produce physically realistic slip models consistent with *Mai and Beroza* [2002].

Where the model space is limited by the fractal constraint on slip values, only a relatively small number of solutions from the remaining subset of models will satisfy the displacement data. A Bayesian Monte Carlo approach to the problem [*Nic Bhloscaidh et al*., 2015] must sample billions of models before it converges on a robust solution. Even then, the Bayesian Monte Carlo technique is sensitive to a number of parameters with tight constraints on the acceptable model size and magnitude necessary to produce timely solutions. The GASE technique relies on a genetic algorithm both to accelerate the identification of models that satisfy the displacement data and to reduce its sensitivity to the initial model parameter selection.

2.1.3. The Genetic Algorithm

The genetic algorithm recombines the information held in a population of models to identify a superior set of models using Darwinian operators and the principles of natural selection [*Whitley*, 1994; *Goldberg*, 1989; *Engelbrecht*, 2002]. Parameters describing model solutions (i.e., the parameters determining slip values and location) are encoded as a set of genes, which we call a model genome. Following this biological analogy,

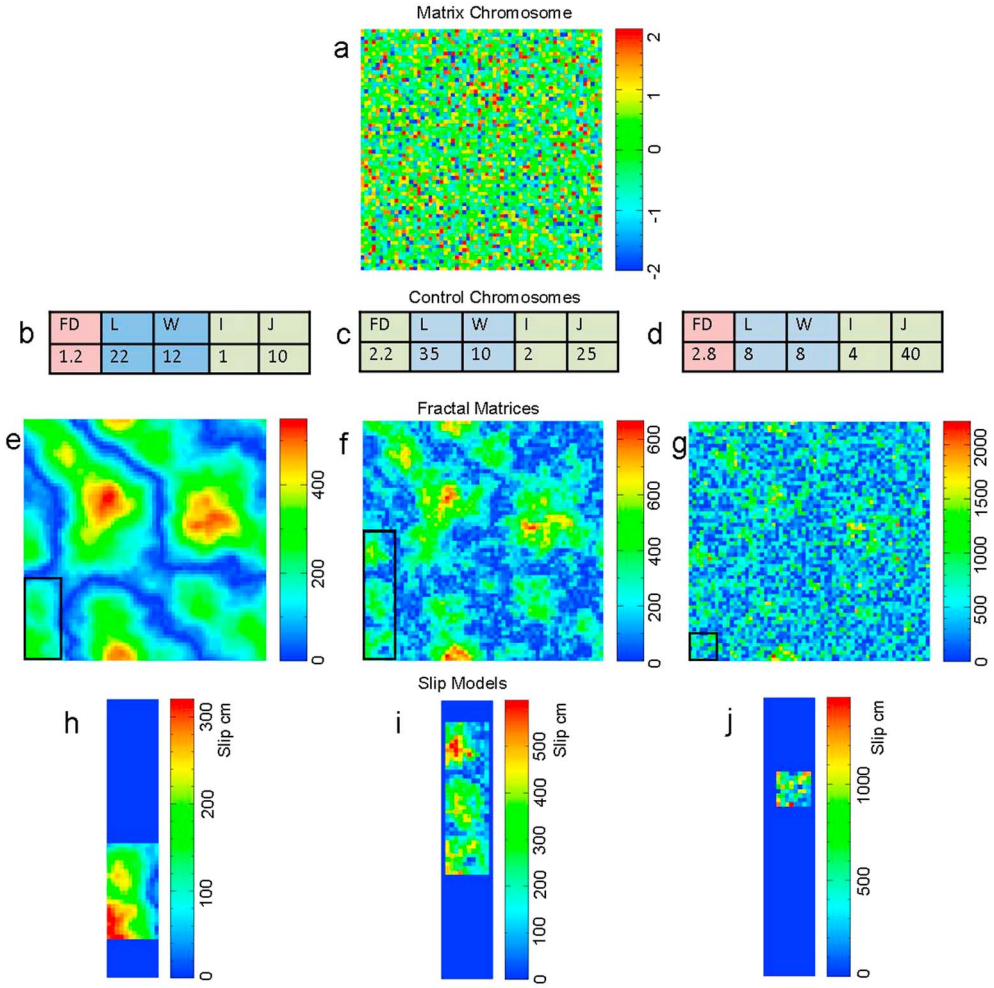

Figure 3. Different configurations of the control chromosome parameters producing different slip models from a single matrix chromosome. Parameters used to describe potential solutions are encoded into a composite data structure consisting of (a) the matrix chromosome, a normalized Gaussian array ($\mu = 0.0$, $\sigma = 1$); (b–d) control chromosome, whose values are used to convert the matrix chromosome to a fractal matrix: FD determines the fractal dimension of the model, L and W limit the size of the model (shown as a black box), and I and J map the model slip distribution onto the fault model; (e–g) how the Gaussian matrix is converted to a fractal matrix; (h–j) three possible realizations of the matrix chromosome using different values on the control chromosome genes. Note the different scales on the fractals and slip models. The maximum slip values expressed on the fault model are determined by the control chromosome genes.

the model phenotype, i.e., the physical appearance and behavior of a solution, refers to the realization of its genome [*Whitley*, 1994; *Goldberg*, 1989; *Engelbrecht*, 2002]. In relation to the earthquake problem, the model phenotype refers to the location and magnitude of slip values. Genes can hold different values and, hence, produce different model phenotypes; for example, varying the values of genes that control the event location causes solutions to be mapped onto different areas of the fault model (Figure 3). Our approach differs from the canonical genetic algorithm schema proposed by *Holland* [1975] which encodes solution parameters as a single, binary string. Instead, solution parameters are encoded as continuous, real values [*Michalewicz*, 1996; *Bessaou and Siarry*, 2001; *Tsai et al*., 2009] and stored in a composite data structure consisting of two parts: a matrix chromosome and a control chromosome. Our technique requires us to produce a number of these model genomes structured to the following specifications:

1. The matrix chromosome: a 64×64 element matrix filled with pseudorandomly generated values drawn from a normalized Gaussian distribution (mean of 0.0 and standard deviation ± 1.0). Following *Turcotte* [1997], we take the Fourier transform of the matrix chromosome and apply a low-pass filter (scaled to the value of the fractal dimension gene, described below) to the Fourier coefficients. We then compute the inverse Fourier transform

of the filtered coefficients to create a fractal matrix. The conversion of the matrix chromosome to the fractal matrix is computationally expensive; hence, we store the fractal matrix. As a consequence, we are able to create multiple slip models by varying the value of genes on the control chromosome (Figure 3a).

2. The control chromosome: a second set of genes that performs a regulatory function controlling the activation and expression of genes from the matrix chromosome [*Dasgupta and McGregor*, 1993; *Tsai et al*., 2009; *Wallet et al*., 1996]. Different configurations of the control chromosome genes are shown in Figures 3b–3d. The genes on the control chromosome convert the information on the matrix chromosome into slip models whose fitness is then evaluated by the algorithm. By altering the value of the control chromosome genes, we can produce multiple realizations from one matrix chromosome (Figures 3e–3g). Genes on the control chromosome have a number of roles:

    a. Gene FD determines the fractal dimension, which is used as a metric for slip correlation, and regulates the gradient of slip between neighboring elements of a solution. Its value is randomly selected from a range of 2.0 to 2.5, consistent with *Mai and Beroza* [2002].
    b. Genes L and W define the size of the model slip distributions copied from the larger fractal matrix, where the bottom left cell of the fractal corresponds to the bottom left cell of the model slip distribution. The L-gene determines the along-strike length, and the W-gene determines the downdip width of the model in kilometers. Each is randomly assigned a value given $120 \leq L \leq j_{\text{len}}$ and $120 \leq W \leq i_{\text{len}}$, where $j_{\text{len}} = 700$ km and $i_{\text{len}} = 240$ km are the along-strike and downdip limits of the slip model, respectively. This equates to an along-strike range of 6 to 35 cells and a downdip range of 6 to12 cells of the fault model.
    c. Genes I and J store the coordinates used to map the lower left corner of the solution onto the fault model. The along-strike coordinate is stored by J-gene and the along dip by the I-gene. Each holds a randomly sampled value from within the legal bounds of the fault plane, so that $1 \leq J \leq (j_{\text{max}} - L)$ and $1 \leq I \leq (i_{\text{max}} - W)$, where $j_{\text{max}} = 1280$ km and $i_{\text{max}} = 240$ km are the along-strike and downdip bounds of the fault plane model, respectively.

2.2. Creation of Synthetic Displacement Sets

It is important that GASE solutions reproduce both the displacement data and features of the causative slip distributions. Therefore, we conduct a series of tests on synthetic data to assess the effectiveness of our technique. Synthetic slip models are created and mapped onto the fault model allowing synthetic displacement sets to be calculated. These synthetic displacement sets are then used by the algorithm to estimate the original synthetic slip model. Knowledge of the details of the input slip model allows straightforward comparisons to the obtained GASE solutions.

A randomly generated 64×64 cell Gaussian matrix is used to produce a two-dimensional fractal matrix (fractal dimension of 2.3). We identify three smaller sections of the fractal matrix so that each fits into the fault model and is not bounded by high slip along its outer edges. We call them synthetic slip models 1–3 (SSM 1–3) (Figures 2e–2g). Furthermore, we create three sets of observation locations: Synthetic Grid 231 (SG231), 025 (SG025), and 015 (SG015) (Figures 2b–2d) and calculate their displacement matrices. The displacements produced by each of the synthetic slip models on each synthetic grid are recorded, giving a total of nine displacement sets. We refer to each displacement set first by the number of observation points in the set, then by the synthetic slip model number. For example, we name the displacement set of 15 observation points which we use to model synthetic slip model 1 as SG-015 m1. Accordingly, the nine displacement sets produced from our three observation grids and three synthetic slip models are referred to as SG-231 m1, SG-231 m2, SG-231 m3, SG-025 m1, SG-025 m2, SG-025 m3, SG-015 m1, SG-15 m2, and SG-015 m3.

To approximate the error structure of the published coral data, artificial uncertainties are added to each observation. For each observation, an error factor is drawn at random from a normalized Gaussian distribution. The uncertainty $\pm\gamma$, associated with each observation point, is given by the product of the synthetic displacement μ and the error factor ϵ:

$$\gamma = \mu \times \epsilon \tag{5}$$

2.3. Creation of a Parent Population

We randomly generate a collection of 100 model genomes according to the specifications described in the previous section, each one consisting of a matrix chromosome and control chromosome. We refer to this

set of models as a parent population. Our final slip estimate relies on the information sampled from 100 of these parent populations, but we consider each population to be independent of one another (We have tested different sizes of parent populations in order to find the optimal size and included the results of these tests in the supporting information.). We recombine the genes held in the parent population to find models better converged on the displacement data. Our parent population must be large enough to hold sufficient information necessary to identify a set of optimal solutions. Moreover, it is important that this initial set of models samples an adequately large area of the model space. This helps prevent the algorithm from prematurely converging on locally optimal solutions, which are of a limited use in data modeling. To ensure that our initial parent population is suitably diverse, we test each newly created genome based on its eligibility to join the population. We set a Hamming Distance to define the minimum acceptable dissimilarity between the values stored in two matrix chromosomes [*Guo and Zhao*, 2002]. As each randomly generated genome is created, we determine the similarity of the new genome's Matrix Chromosome and the matrix chromosomes of the genomes present in the initial population by calculating the geometric distance d defined as

$$d_{al} = \sum_{ij} \left(M_{ij}^{a} - M_{ij}^{l} \right)^2 \tag{6}$$

where M_{ij}^{a} refers to an element ij of the matrix chromosome a, and M_{ij}^{l} refers to an element of the matrix chromosome l in the parent population. As each new genome is accepted into the parent population, the current parent population size N_{cur} increases by 1, $l = 1, N_{\mathrm{cur}}$ until we reach the target population size of 100 models (the results of tests determining the optimal population size have been included in the supporting information). If $d_{al} < H$, then M^{a} is rejected from the parent population and a new genome is pseudorandomly created. Model creation continues until the initial population reaches the size N_{cur}. Through trial and error we identified $H = 4300$ as a point above which the algorithm had difficulty finding sufficient models to fill the initial population for a 64×64 cell matrix. The result is a quasi-random population distributed throughout the model space. After establishing the initial population, the genomes are converted to their phenotypic forms and are fitness tested.

2.4. Evolution of the Parent Population

We compute the fitness of each of the 100 models in the parent population and then recombine the information stored in each model genome, i.e., information on the matrix chromosome and control chromosome, to produce new models using the following three operators: selection, crossover, and mutation.

1. Selection determines how two models are chosen from the parent population to undergo crossover (as described below) [*Whitley*, 1994; *Goldberg*, 1989; *Engelbrecht*, 2002]. We have tested a number of different selection techniques and determined "BottomUp" to be the most successful (see the supporting information). Accordingly, we choose each model from the parent population in ascending order of fitness (i.e., their ability to reproduce the displacement data), then pair it with a second randomly selected model of equal or lower fitness.
2. Crossover allows the information stored in the selected parent models (i.e., the genes of the matrix chromosome) to be recombined to produce a series of new offspring models [*Whitley*, 1994; *Goldberg*, 1989; *Engelbrecht*, 2002]. Values are randomly copied from the matrix chromosome of either parent model to form the matrix chromosome of an offspring model. The greater the number of offspring models, the greater the likelihood of finding models that better satisfy the displacement data. However, producing more offspring models increases the computational costs associated with the technique. Therefore, using synthetic data, we have determined the optimal number of offspring models, produced during each crossover operation, to be 50 (see supporting information). We have also tested a number of different crossover implementations (see supporting information) and determined "Cluster4×4" technique to be the most effective. Here genes from the parent matrix chromosome are copied to the offspring matrix chromosome in 4×4 cell clusters. Simultaneously, genes from the control chromosomes of either of the parent models are copied at random to the offspring model. We use Roulette Wheel Sampling to bias the transfer of information in favor of the fitter parent [*Whitley*, 1994; *Engelbrecht*, 2002]. The value R^* is given by

$$R^* = 1 \left(1 + \frac{F_2}{F_1} \right) \tag{7}$$

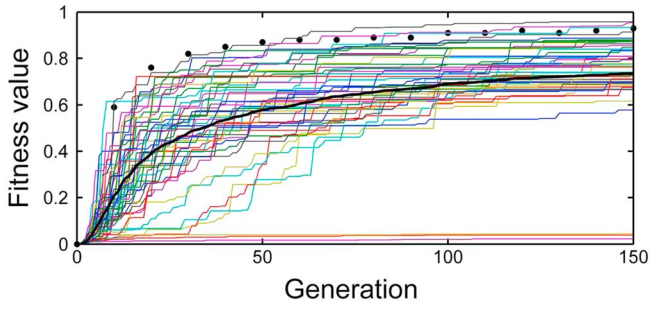

Figure 4. Changes to model fitness over 150 generations. The fitness of the best model in 50 parent populations (colored lines) are shown alongside the average model fitness (black line). An Ensemble Estimate model (as described in section 2.5) is produced every 10 generations and its fitness plotted as a series of black points.

where F_1 is the fitness of Parent$_1$ and F_2 is the fitness of Parent$_2$. A random number R_1 between 0 and 1 is generated, where $R^* \leq R_1$, Parent$_1$ contributes information to the offspring model and where $R^* > R_1$, Parent$_2$ contributes information to the offspring model.

3. Mutation allows the values of genes on the matrix chromosome to be altered as they are copied from parent to offspring solutions. This in turn affects the values of the fractal matrix representing our earthquake slip. Where the algorithm has converged on local optima, mutation adds new information to the solution, potentially opening up new areas of the model space. Mutations occur at random and may destroy information evaluated as being useful so its effectiveness is governed by the size and frequency of the mutation [*Whitley*, 1994; *Goldberg*, 1989; *Engelbrecht*, 2002]. Our synthetic tests (see supporting information) show that the effects of mutation are generally small and fail to produce a consistent positive effect on solutions. As a result, we have chosen not to use a mutation operator.

After each selection/crossover operation the offspring models are converted to their phenotypic form (i.e., mapped as a slip distribution onto the fault model) and are fitness tested. Where the offspring models are found to be fitter than the selected parent models, they replace them so that the two fittest models are returned to the parent population. Note that the size of the parent population remains constant.

We refer to a complete set of selection/crossover operations as one generation (i.e., where each model has been selected in turn and paired with another to undergo crossover). As the number of generations increases, so does the fitness of models in the parent population (Figure 4). Systematic tests have shown that improvement in model fitness occurs most rapidly over the first 50 generations (see supporting information); however, we suggest the number of generations be set to 100, as above this point improvements to fitness generally plateau.

Figure 5 shows the displacement sets produced by the initial parent population of 100 models and by stack models from 100 parent populations. When comparing the displacements of the initial population to the target displacements SG015m1, we observe that the displacements produced at each observation point fall mostly outside the uncertainty range of the target displacements (Figure 5a). By comparison, the displacement sets produced by the stack models that have evolved over 100 generations are closely converged on the target displacement set and in most cases fall within the given uncertainties for each observation (Figure 5b).

2.5. Ensemble Model

As previously stated, multiple models are capable of reproducing the sparse data so our forward modeling approach identifies a number of models consistent with the observations. We create a number of parent populations and allow each to evolve independently. After 100 generations, we identify the best model from each population and copy it to a catalog to which we refer to as a stack [*Guo and Zhao*, 2002]. We weight each stack model by its fitness, and then, from the stack, we generate an Ensemble Estimate model (EE model) of the slip distribution. As the fitness of each individual model is purely determined by its ability to reproduce the displacements for a given set of observation points, those models that produce a good fit to the data have the strongest contribution to the EE model slip. We calculate the EE model so

$$\mathrm{EE} = \left(\frac{1}{\sum_i^N F_i}\right) \sum_i^N M_i . F_i \tag{8}$$

where M_i is each model present in the stack, F_i refers to the corresponding fitness value, and N is the total number of models selected by the algorithm (Figure 6). Systematic tests have shown that a robust EE model

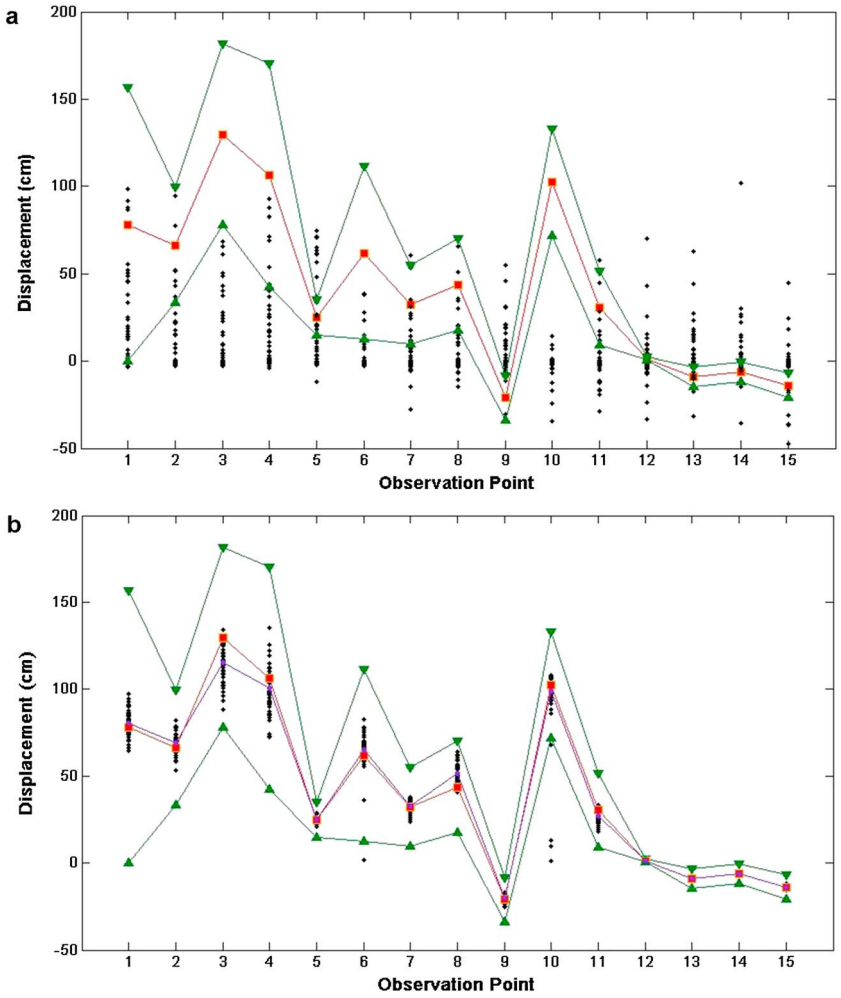

Figure 5. The displacement sets produced by (a) the initial population of 100 models and (b) 100 stack models plotted as black dots. The target displacements SG015m1 are shown in red along with its uncertainties in green. The displacements generated by the Ensemble Estimate model (as described in section 2.5) are shown in purple.

can be produced by sampling the best model from each of 100 parent populations (see supporting information). In Figure 4 we examine how the fitness of the EE model changes as the number of generations increases. We observe high variability in the stack model fitness, with only three models outperforming the EE model. Furthermore, while the EE model produces a displacement set closely converged on the target displacement set, some of the stack models fail to reproduce uplift within the given uncertainties.

3. Testing the Genetic Algorithm Slip Estimator

Using synthetic data described in section 2.2, we have carried out extensive tests of the GASE technique examining the implementation of the genetic algorithm operators and optimization of its parameters, i.e., the size of the parent population, the number of offspring produced during crossover, the number of generations, and the number of populations. We have included the details of these tests and results in the supporting information. Here we analyze the sensitivity of generated slip estimates to the initial parent population constraints. We go on to examine the EE models produced for 12 synthetic slip distributions using the optimal settings derived from the synthetic tests.

3.1. Evaluation of the EE models

In addition to evaluating the EE model fitness, we examine the similarity of our slip estimates to the corresponding input slip model and examine the variability of slip values in the stack models. Based on the City

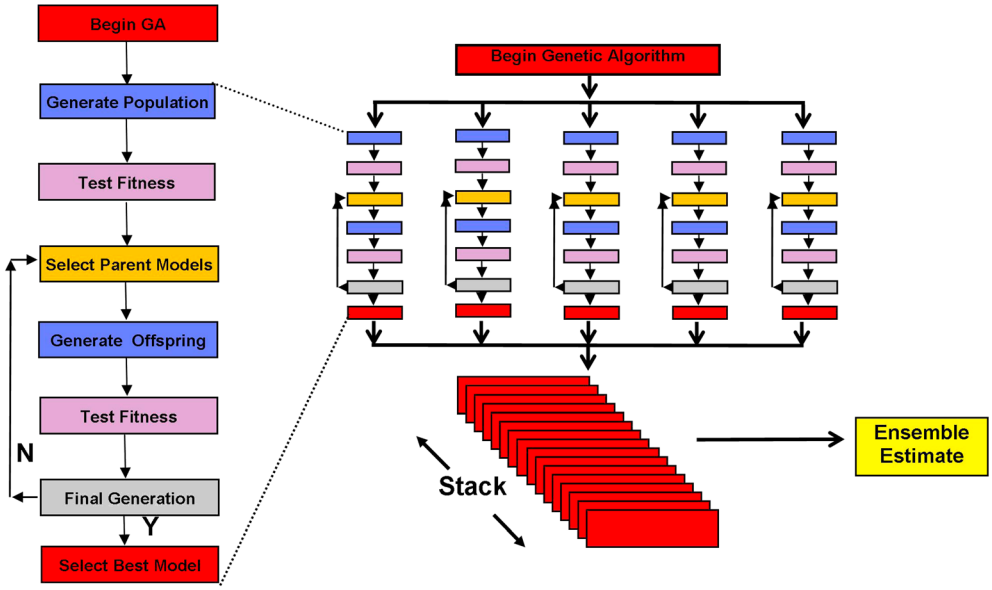

Figure 6. Flowchart outlining the GASE algorithm. The technique generates a number of parent populations. The genetic algorithm recombines the information in each population until it converges on a set of solutions satisfying the fitness function, in this case, reproducing a set of displacements. The best model from each population is stored and stacked. Weighted by their fitness, the stack models are used to create an ensemble estimate of the slip distribution responsible for the given displacements.

Block Dissimilarity [*Webb*, 2002], we calculate the e value as the absolute difference between values in corresponding cells of the target synthetic slip model, t_{ij} and EE model, v_{ij}, normalized to the peak slip value of the target synthetic slip model $t_{\max}$ so

$$e_{ij} = \left| \frac{v_{ij} - t_{ij}}{t_{\max}} \right| \tag{9}$$

The e values are summed for each cell of the solution and divided by the total number of cells in the fault model N_c to give a normalized average absolute dissimilarity (E value) for each EE model:

$$E = \frac{1}{N_c} \sum\nolimits_{ij} e_{ij} \tag{10}$$

We also calculate the normalized average absolute deviation to examine the variability of slip values in the stack of models used to compute the ensemble estimate solution. Treating slip in each cell of the EE model v_{ij}, as a measure of central tendency, the City Block Distance is calculated for the corresponding cell in each of the stacked models M_{ijq}, where q is the total number of sampled population [*Webb*, 2002]. We normalize each value by the peak slip in the EE model $v_{\max}$, and in effect, we obtain s value reflecting the average deviation of slip in corresponding cells of the stack models:

$$s_{ij} = \frac{1}{N_q} \sum\nolimits_{q} \left| \frac{M_{ijq} - V_{ij}}{v_{\max}} \right| \tag{11}$$

Summing the s values for each cell and averaging over the total number of cells gives an S value for the stack:

$$S = \frac{1}{N_c} \sum\nolimits_{ij} s_{ij} \tag{12}$$

We also calculate the moment magnitude (M_w) of EE models and compare it to the M_w of the target slip distribution. The earthquake moment M_o is defined as

$$M_o = \mu A D \tag{13}$$

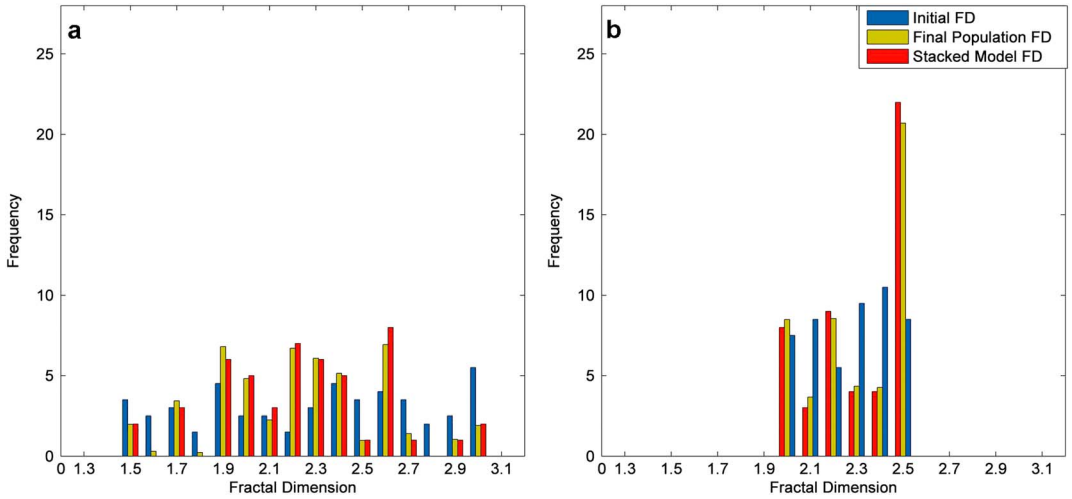

Figure 7. Comparing the fractal dimension values of models in the initial and final parent populations and in the catalog of stack models. We examine the distribution of the fractal dimension values in the initial population (blue), the final population (yellow), and the models cataloged for the stack (red) observed for (a) the FreeFD scenario (values randomly chosen from the range 1.5 to 3.0) and (b) the LimitedFD scenario (values randomly drawn from the range 2.0 to 2.5). As only one model from each final population is chosen for the stack we divide the frequency values of the initial and final parent populations by the total population size (100 models).

where μ is the shear modulus (33 GPA); A is the area (m^2), and D is the summed slip (m). From this the moment magnitude (M_w) is calculated:

$$M_w = \frac{2}{3}\log 10(M_o) - 6. \tag{14}$$

3.2. Parent Population Properties

Source parameters (e.g., locations and event extent) for recent earthquakes are well defined. For events in the last few centuries, where available, written records can complement and constrain paleoseismic observations. Where existing knowledge can inform the modeling process, it should be used. However, extending the paleoseismic record even further back relies on more deficient observational data and an absence of written records. By examining the value of the parameters constraining the generation of models in the initial parent population, we can test the algorithm's ability to recover details of poorly constrained rupture patterns.

3.2.1. Testing Fractal Dimensions

We test a number of scenarios to examine how different ranges of the fractal dimension in the initial parent population impact the final EE model: (a) FixedFD, the value of the fractal dimension for each model of the initial population is set to the fractal dimension of the target slip model; (b) LimitedFD, the fractal dimension for each model of the initial population is randomly drawn from the range 2.0 to 2.5; and (c) FreeFD, the fractal dimension for each model of the initial population is randomly chosen from the range 1.5 to 3.0. We test the different fractal dimension scenarios using the LimitedES constraint on model size, as described below. The EE models are produced, for each of the nine displacement sets described in section 2.2, and their fitness is evaluated. The solution fitness, normalized by the number of observation points in the corresponding displacement set, is computed and ranked to allow the impact of specific fractal dimension values to be examined. In Figure 7 we compare the distribution of fractal dimension values in the initial parent population to those in the final parent population and the catalog of stack models for the LimitedFD and FreeFD scenarios.

For the solutions to each set of displacement data, we assign a rank from 1 to 3 to each scenario based on the EE model fitness. For example, the normalized model fitness for SG-015 m2 is 0.992 for FixedFD, 0.987 for LimitedFD, and 0.989 for FreeFD. Therefore, the corresponding ranks are 1, 3, and 2, respectively. Having ranked each solution, we sum the ranking values for each scenario to obtain its total rank value R. The lowest value of the total rank R is given the best final rank, FR = 1, whereas the highest value of the total rank R is assigned FR = 3. We observe that the solutions are generally better ranking (i.e., where FR is lowest) when we restrict the initial range of fractal dimension values (Table 1). For the FreeFD selection method, the EE

Table 1. EE Model Fitness Values for Different Range Constraints of the Fractal Dimension[a]

	SG-015			SG-025			SG-231				
	SSM 1	SSM 2	SSM 3	SSM 1	SSM 2	SSM 3	SSM 1	SSM 2	SSM 3	*R*	FR
FixedFD	0.990$^{(2)}$	0.988$^{(2)}$	0.999$^{(1)}$	0.987$^{(1)}$	0.977$^{(1)}$	0.995$^{(1)}$	0.308$^{(1)}$	0.404$^{(1)}$	0.372$^{(1)}$	11	1
LimitedFD	0.992$^{(1)}$	0.990$^{(1)}$	0.999$^{(1)}$	0.986$^{(2)}$	0.972$^{(2)}$	0.994$^{(2)}$	0.256$^{(2)}$	0.398$^{(3)}$	0.353$^{(2)}$	16	2
FreeFD	0.989$^{(3)}$	0.986$^{(3)}$	0.999$^{(1)}$	0.983$^{(3)}$	0.968$^{(3)}$	0.993$^{(3)}$	0.233$^{(3)}$	0.400$^{(2)}$	0.317$^{(3)}$	24	3

[a]Fitness values are normalized by the total number of observation points given nine different synthetic displacement sets. Numbers in brackets are ranks for individual solutions. *R* is the total rank assigned to each selection method (FixedFD, LimitedFD, and FreeFD). The value of FR, ranging from 1 to 3, represents the final rank where 1 refers to the most preferable and 3 to the least preferable method.

models still reproduce the main slip features of the target synthetic slip model, although the along-strike extent of the solutions is poorly defined (Figure 8).

We also compare the distribution of fractal dimension values present in the (1) initial population, (2) final population, and (3) stack models (Figure 7). For both FreeFD and LimitedFD selection methods, the observed frequency distribution of fractal dimension values in the final population closely matches that of the stack models (Figure 7). For FreeFD, 82% of the fractal dimension values of the stack models are nonuniformly distributed between 1.9 and 2.6 and 54% between 2.0 and 2.5; however, we report no convergence on a specific value (Figure 7a). Where the fractal dimension is further constrained to the LimitedFD range, the values in the final population and the stack models concentrate at the upper limit of 2.5 (Figure 7b).

3.2.2. Slip Model Size

In order to evaluate the effect of the model size on the EE model fitness, we test three different scenarios: (a) FixedES, the model size is the same as the size of the target synthetic slip model; (b) LimitedES, the along-strike model length ranges from 400 to 800 km, and the downdip width ranges from 160 to 240 km, the equivalent of 20 to 40 cells along and 8 to 12 cells across the fault model; and (c) FreeES, the along-strike model length ranges from 120 to 1200 km, and the downdip width ranges from 120 to 240 km, the equivalent of 6 to 60 cells along and 6 to 12 cells across the fault model. As the fractal dimension tests demonstrate that the LimitedFD scenario yields the fittest EE models, we apply the LimitedFD constraint to models in the initial parent population to test different model scenarios. In each case, solutions from the nine displacement sets (see section 2.2) are produced and their normalized fitness computed. The EE models for each displacement set are ranked in order of fitness to evaluate each of the different model size scenarios (Table 2).

We find that for the FixedES scenario (i.e., when the size of models is known), our solutions have a higher value of normalized fitness. Outside of our synthetic simulations, the dimensions of the earthquakes are unknown, particularly when we are dealing with preinstrumental events, so we will focus on the solutions to the LimitedES and FreeES scenarios. Where the length and width of the initial model are poorly constrained (i.e., LimitedES and FreeES), we still correctly map slip onto the appropriate sections of the fault model; however, we report greater levels of artifact slip along strike (Figure 8). We recover the broad features of the target synthetic slip model, but some of the finer detail is lost, and hence, our solutions are oversmoothened versions of the input slip distribution. For the LimitedES and FreeES scenarios, the EE models tend to underestimate peak slip values, though they successfully reproduce the slip asperities present in the target model. Further from the observation points, our solutions reproduce less detail of the input slip model; however, they do not introduce spurious slip asperities.

Since the EE model is computed from the stack of the fittest models in each parent population, we obtain more information when these individual models are in agreement (i.e., the signal-to-noise ratio is high), and therefore less information where there is disagreement (i.e., the signal-to-noise ratio is low). Where the signal-to-noise ratio is highest, i.e., proximal to the surface displacement, the agreement among the stack models gives us a relatively good estimate of the target slip. Despite this, the small variations in the slip values of the stack models produce the smoothed-out appearance of our final EE models. We report greater disagreement among the stack models further from the observation points resulting in poor estimation of slip values. In the case of FixedES and FreeES scenarios, where we have poorly constrained length and width values, we observe slip extending along the length of the fault model. Away from the observation points,

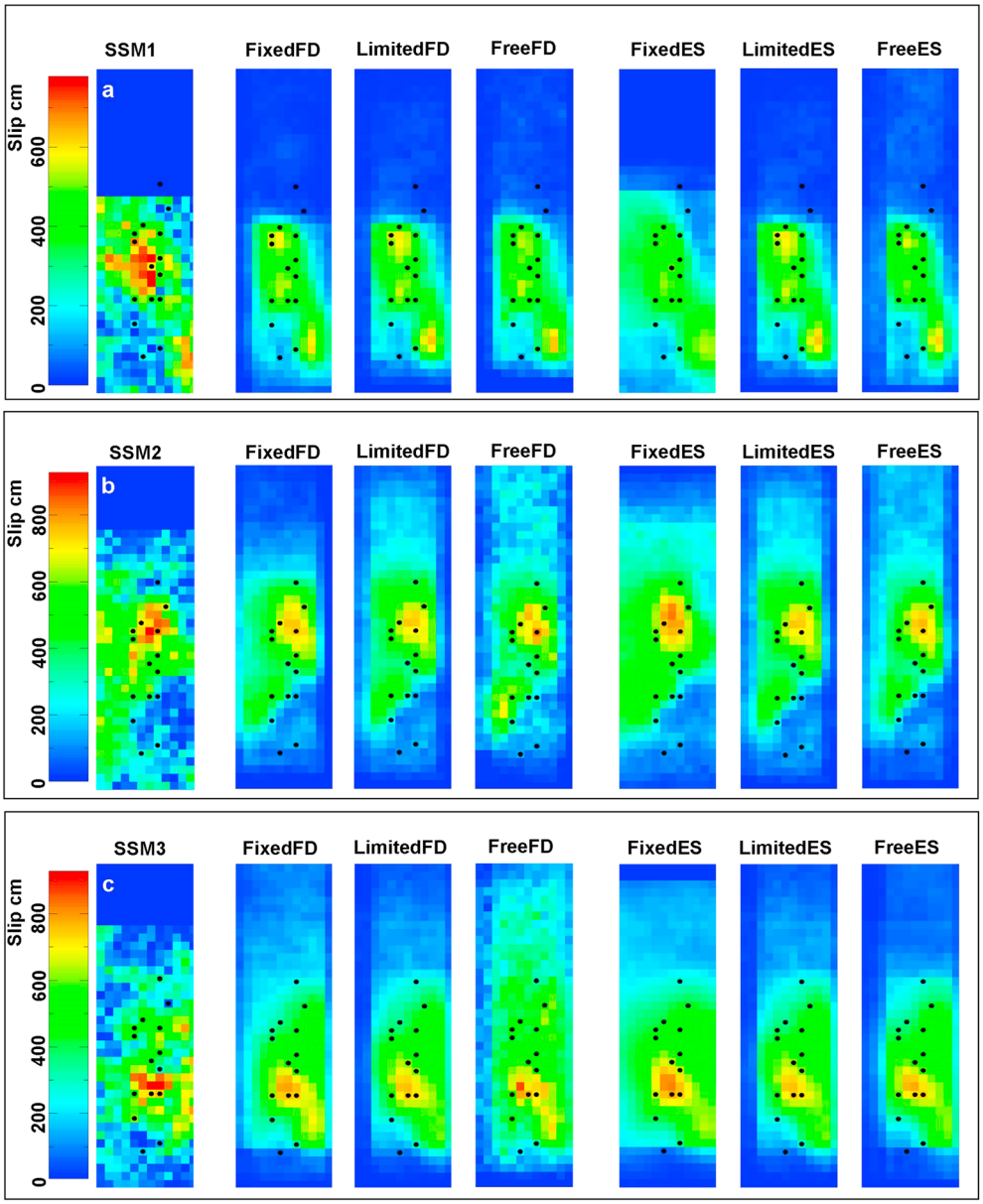

Figure 8. Examining the sensitivity of EE models to different initial parent population and fractal dimension constraints. We compare the EE models produced from the different fractal dimension and model size scenarios described in sections 3.2.1 and 3.2.2 for (a) synthetic slip model 1, (b) synthetic slip model 2, and (c) synthetic slip model 3. The LimitedES constraints were applied to the initial population for our tests of the different fractal dimension scenarios. We applied the LimitedFD constraints to the initial population when examining the different model size scenarios.

Table 2. Normalized EE Model Fitness for Different Ways of Constraining the Initial Size of the Model[a]

	SG-015			SG-025			SG-231				
	SSM 1	SSM 2	SSM 3	SSM 1	SSM 2	SSM 3	SSM 1	SSM 2	SSM 3	*R*	FR
FixedES	$0.982^{(2)}$	$0.991^{(1)}$	$0.998^{(2)}$	$0.986^{(1)}$	$0.976^{(1)}$	$0.993^{(2)}$	$0.933^{(1)}$	$0.628^{(1)}$	$0.931^{(1)}$	12	1
LimitedES	$0.983^{(1)}$	$0.987^{(2)}$	$0.999^{(1)}$	$0.984^{(2)}$	$0.970^{(2)}$	$0.994^{(1)}$	$0.360^{(3)}$	$0.399^{(2)}$	$0.309^{(3)}$	17	2
FreeES	$0.983^{(1)}$	$0.985^{(3)}$	$0.999^{(1)}$	$0.983^{(3)}$	$0.962^{(3)}$	$0.994^{(1)}$	$0.455^{(2)}$	$0.136^{(3)}$	$0.410^{(2)}$	19	3

[a]Fitness values are normalized by the total number of observation points given nine different synthetic displacement sets. Numbers in brackets are ranks for individual solutions. *R* is the total rank for a given selection method (FixedES, LimitedES, FreeES). The value of FR, ranging from 1 to 3, represents the final rank where 1 refers to the most preferable and 3 to the least preferable method.

greater variability in possible slip values gives a lower signal-to-noise ratio. Consequently, we observe low levels of artifact slip extending along strike and around the periphery of our EE model. While individual stacked solutions may have slip asperities along their edges or beyond the visibility of the observation points, the contribution of uncommon stack features to the EE model is small.

3.3. Synthetic Displacements on Actual Coral Locations

To determine how well we can estimate slip distributions from real coral locations, we attempt to recover features of the synthetic slip models from a set of 11 observation points analogous to the 1797 Mentawai earthquake coral displacement set [*Natawidjaja et al*., 2006] (Figure 2h). Following the procedure described in section 2.1.1, we create a discretized fault model and a displacement matrix for each coral location. We set the GASE parameters to the optimal values in Table 3, as identified by systematic testing (see supporting information). To further test the robustness of the technique, we generate an additional nine synthetic slip models following the procedure described in section 2.2. Each of these new synthetic slip models is selected so that we have a range of different sized events. Note that we shift the position of SSM 1–3 so the slip lies on the fault plane under the locations of observation points. We produce a displacement set for each of the 12 synthetic slip models and assign uncertainties proportional to the published errors to each observation point [*Natawidjaja et al*., 2006]. We then apply the GASE algorithm to the data and for each compute an EE model.

The slip values of the resulting EE models are of similar range to the target model and are in most cases confined to the correct section of the fault model (Figure 9); however, the maximum slip values of the EE models are generally lower than those of the corresponding target models. While we are able to reproduce the broad features of the target slip distributions, some of the finer details of the input slip distribution are lost. Our solutions are generally well constrained along strike. At the same time, we observe that the slip values in the downdip direction from the observation points are better estimated than those updip. All EE models satisfy the target displacement data within the given uncertainties. In 50% of cases, our EE models are within M_w ±0.05 of the target synthetic slip models and each of our remaining solutions are within M_w ±0.12 of the target models (Table 4). We have previously observed that slip is generally well estimated close to the observation points where the signal-to-noise ratio in stack models is low. Further from the observation points, the signal-to-noise ratio in stack models decreases and we report oversmoothening our modeled slip. Where there is insufficient displacement data to constrain the along-strike and along-dip limits of our solutions, declining levels of slip extend along the fault plane causing us to overpredict the M_w of the target models. For example, our solutions to synthetic slip models m006 and m007 overestimate the target M_w by 0.07 and 0.11, respectively. Conversely, in some cases we fail to identify features of the target slip distribution along sections of the fault plane unresolvable to the observation points, leading us to underestimate the M_w of the target model. For example, we are unable to correctly identify a slip asperity downdip of the observation points for synthetic slip model m012 and a slip asperity updip of the observations for synthetic slip model m005. In both cases this leads to an underprediction of the target M_w by 0.12 and 0.08, respectively. Therefore, having only limited data, such as in a case of paleoearthquakes, we may expect our technique to give an estimate of the event M_w, rather than determine its precise value.

4. Modeling Real Displacement Data

We now apply our technique to the coral displacement record for the 1797, 1833, 2005, and 2007 earthquakes and evaluate the resulting slip distribution for each event. First, we produce models of the 2005 and 2007 events using only the published coral displacement data [*Konca et al*., 2007, 2008]. This allows us to compare our EE solutions to the previously published slip inversions and gain insight into the potential limitations of estimating paleoslip from the coral data. Next, we compute slip estimates for the 1797 and 1833 paleoearthquakes using the coral data by *Natawidjaja et al*. [2006]. Note that the recent study by *Philibosian et al*. [2014] uses the expanded set of coral observation points to produce the slip distributions of these two paleoearthquakes. We plan to rigorously analyze the supplemented uplift data in the future work.

4.1. Fault Plane Geometry

Following the procedure described in section 2.1, a fault model with an origin point of 7°S and 102°E, strike of 325°, and dip angle $\delta = 15°$ is created for the 1797, 1833, and 2007 earthquakes, to allow the EE models to be

Table 3. Optimal GASE Parameters Used to Produce EE Models as Determined From the Synthetic Tests[a]

Population Size	Number of Generations	Number of Populations	Number of Offspring	Range of FD	Length		Width	
					Cells	km	Cells	km
100	50	100	50	2.0–2.5	10–30	200–600	8–12	160–240

[a]FD refers to the fractal dimension of models in the initial population.

compared with other published models [*Natawidjaja et al*., 2006; *Konca et al*., 2008]. For the 2005 Nias-Simeulue earthquake, *Konca et al*. [2007] tested a number of planar fault models with different angles of dip and found $\delta = 10°$ allowed them to best fit the data. Accordingly, we construct a fault model with a strike of 325° and dip of 10° to be able to assess the EE models against the combined inversion of teleseismic and geodetic data. Here we set the origin point at 6.7°S and 101.8°E to allow the fault model to extend fully under Simeulue Island.

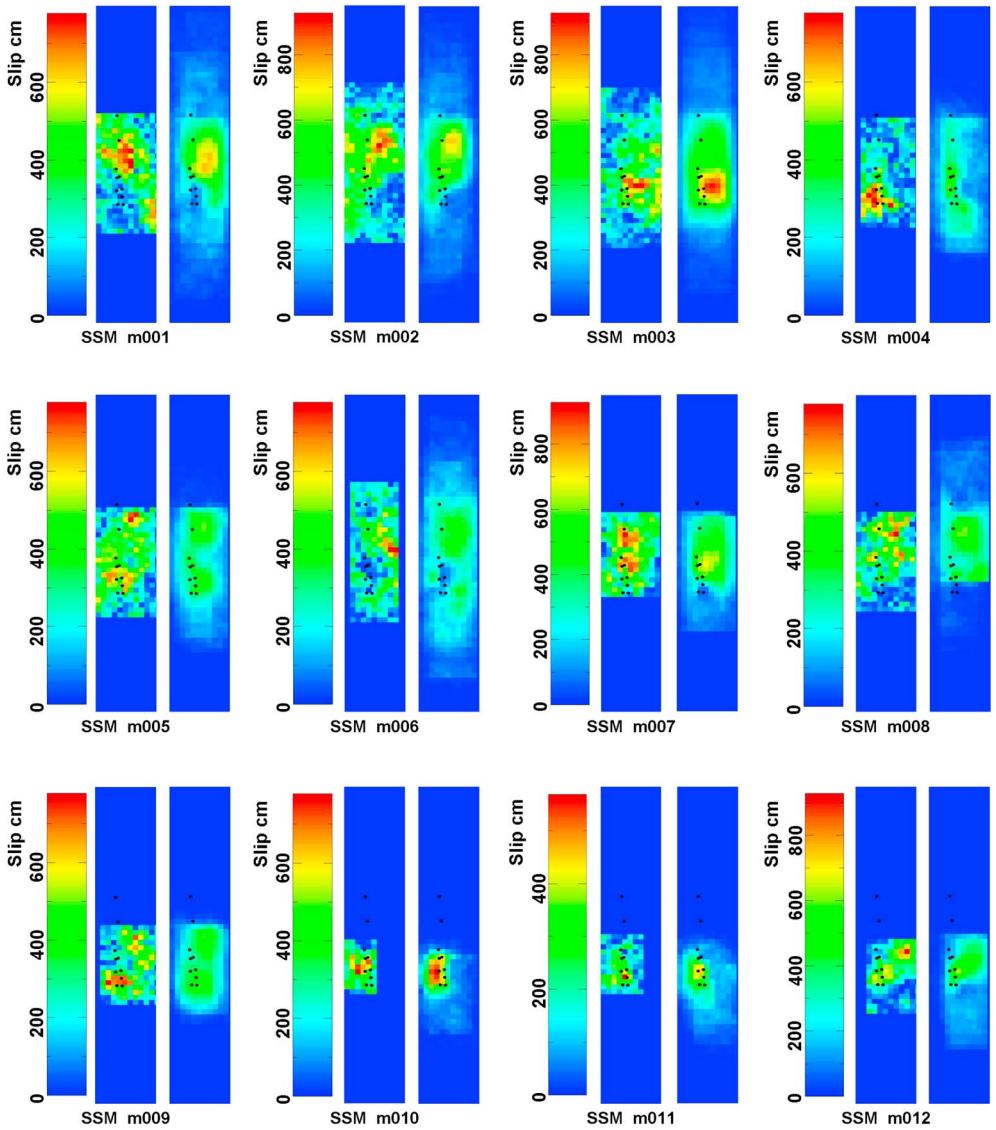

Figure 9. A fault model of the Sunda Trench along with the observation points based on the location of corals uplifted during the 1797 Mentawai earthquake. We use 12 synthetic slip models to generate synthetic displacement sets and then use the GASE technique to estimate the original synthetic slip model. The figure shows the slip distribution of the target models with the corresponding slip distributions for the EE models.

Table 4. Earthquake Magnitude (M_w) Calculated for the 12-Input Synthetic Slip Model and the Corresponding EE Models

	SSM											
	1	2	3	4	5	6	7	8	9	10	11	12
Input M_w	8.72	8.81	8.78	8.51	8.68	8.61	8.66	8.59	8.49	8.28	8.23	8.52
EE M_w	8.71	8.72	8.79	8.52	8.60	8.68	8.61	8.55	8.46	8.35	8.34	8.40

4.2. Modeling the Coral Data for Instrumentally Constrained Earthquakes

4.2.1. The 2005 Nias-Simeulue Earthquake

The March 2005 M_w 8.6 Nias-Simeulue earthquake ruptured the portion of the Sunda megathrust adjacent to the southern extent of the region affected by the 2004 M_w 9.2 Andaman-Sumatran earthquake [*Nalbant et al.*, 2005]. During the earthquake, uplift of up to 2.9 m was recorded by coral heads populating 90 sites along the coastlines of Nias and Simeulue islands [*Briggs et al.*, 2006]. These data were used in conjunction with the uplift data derived from the recently installed campaign GPS (cGPS) arrays to provide the most detailed map of slip distribution [*Briggs et al.*, 2006; *Konca et al.*, 2007].

In Figure 10, we compare our slip estimate to the Kon05 model [*Konca et al.*, 2007]. Both the combined inversion and the EE model have similar along-strike extent. The Kon05 model has peak slip values slightly greater than 12 m and a relatively compact distribution of slip, whereas the EE model produces marginally lower slip values, spread over a wider area. Both the Kon05 and the EE model are of M_w 8.6. The smallest dissimilarities between the two models are along the NE coast of Nias and the southern flank of Simeulue, close to the densest cluster of observation points. Higher *e* values are found off the SW flank of Nias and in the area between the Nias and Simeulue islands where data are sparser. The *S* value showing the average deviation of slip in cells of the stacked models is 0.05 (Table 5). Despite managing to reproduce only 63 of the 90 coral displacements within the observational uncertainty (Figure 12c), the EE model produces similar slip features as the Kon05 model (Figure 10). When we calculate the displacements produced at each observation point by the Kon05 model and compare them to the displacements produced by the EE model, we report that the GASE solution produces a better fit to the displacement data (Figure 12e).

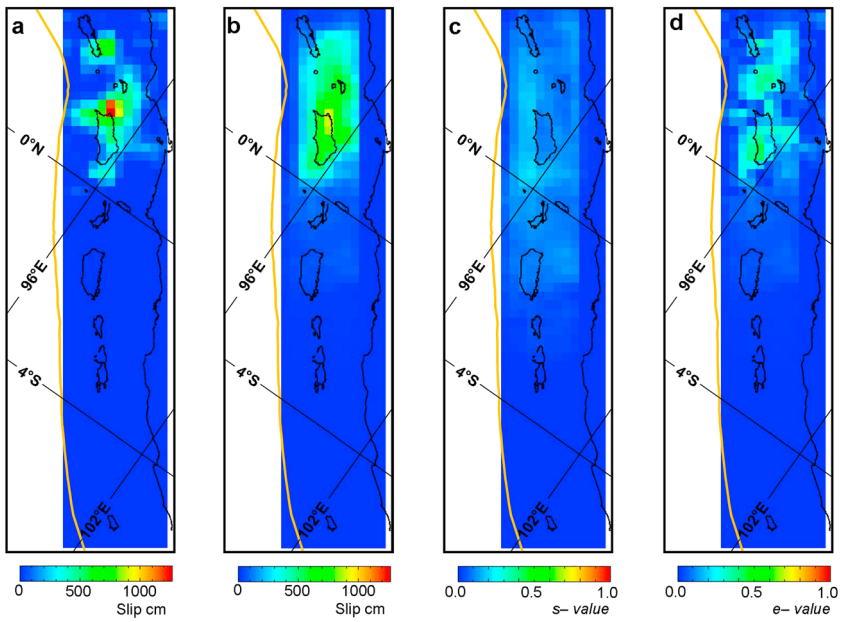

Figure 10. (a) The slip distribution derived from the combined inversion of the teleseismic and geodetic data by *Konca et al.* [2007]; (b) the slip distribution of the EE model; (c) *s* values showing the variability in slip values of the stacked models; and (d) *e* values describing the level of dissimilarity between the EE and Kon05 models calculated for the 2005 Nias-Simeulue earthquake.

Table 5. Comparison of Main Features of the EE Models (the Percentage of Uplift Data Modeled Within Uncertainty, Estimated M_w, and S Value) Calculated for the 1979, 1833, 2005, and 2007 Earthquakes

	1797	1833	2005	2007
% of uplift data modeled within uncertainty	100.0	100.0	70.0	94.4
S value	0.1	0.08	0.05	0.1
M_w	8.7	8.8	8.6	8.7

4.2.2. The 2007 Mentawai Sequence

In September 2007, the Southern Mentawai Islands experienced a sequence of large earthquakes in close succession. The first event, M_w 8.4, began 70 km south of Mega Island and propagated northward to rupture two slip asperities: one, to the north of Mega Island, experienced slip of over 6 m; the other just under the South Pagai produced slip of >7 m. The initial event was followed 12 h later by a second M_w 7.9 earthquake, giving a combined M_w 8.5 [*Konca et al.*, 2008]. The earthquakes displacements for both events were recorded instrumentally by a number of GPS and satellite instruments as well as by 18 coral microatolls [*Konca et al.*, 2008].

We compare our EE model to the Kon07 inversion [*Konca et al.*, 2008] in Figure 11. The along-strike extent of the EE model is poorly constrained compared to the Kon07 model, with low levels of slip extending from the Enggano Island in the south to the north as far as the Batu Islands. Our solution reproduces the distinctive dumbbell-shaped slip asperities observed in the initial M_w 8.4 earthquake, but it fails to clearly identify the later M_w 7.9 event [*Konca et al.*, 2008]. At the same time, it matches the peak slip values of the Kon07 model under Mega Island and South Pagai and successfully models 17 of 18 displacements. The close convergence of the modeled displacements onto the observed uplift can be seen in Figure 12d. A comparison of the displacements produced by both the EE and Kon07 models shows that our solution produces a better fit to the recorded uplift data (Figure 12f). However, the larger area of high slip in the EE model results in a larger M_w of 8.7 when compared to the M_w 8.5 of the Kon07 model.

The greatest dissimilarity between the EE and Kon07 model is indicated by the high e values off the south and west coasts of the Pagai Islands, particularly around Mega Island, where slip is poorly constrained by a few isolated observations. This dissimilarity is reflected in the variability of slip values of the stack models contributing to the EE model. We find much higher s values in the range from 0.4 to 0.5 for the areas around Mega Island than under the Pagai Islands where the s values fall to between 0.1 and 0.2. Better constrained slip in

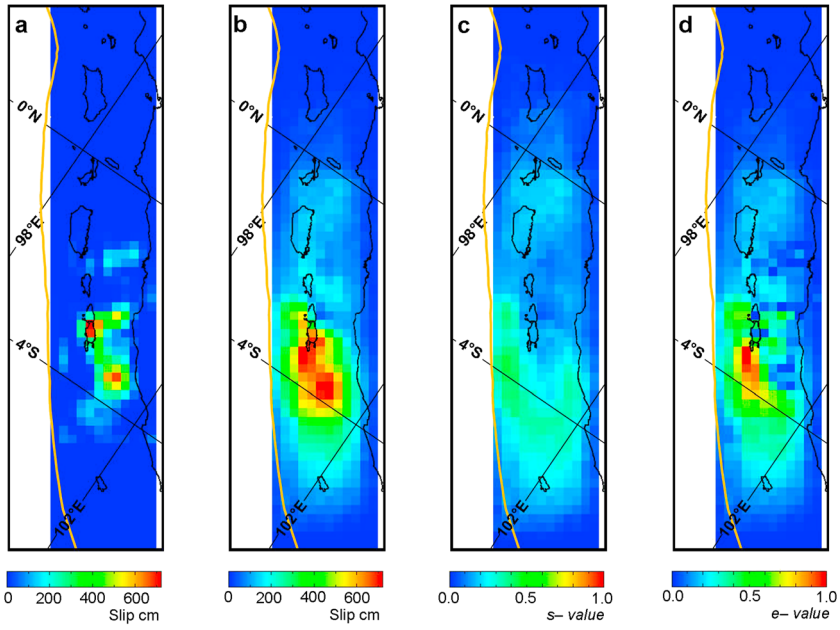

Figure 11. (a) The slip distribution derived from the combined inversion of the teleseismic and geodetic data by *Konca et al.* [2008]; (b) the slip distribution of the EE model; (c) s values showing the variability in slip values of the stacked models; and (d) e values describing the level of dissimilarity between the EE and Kon07 models calculated for the 2007 Nias-Simeulue earthquake.

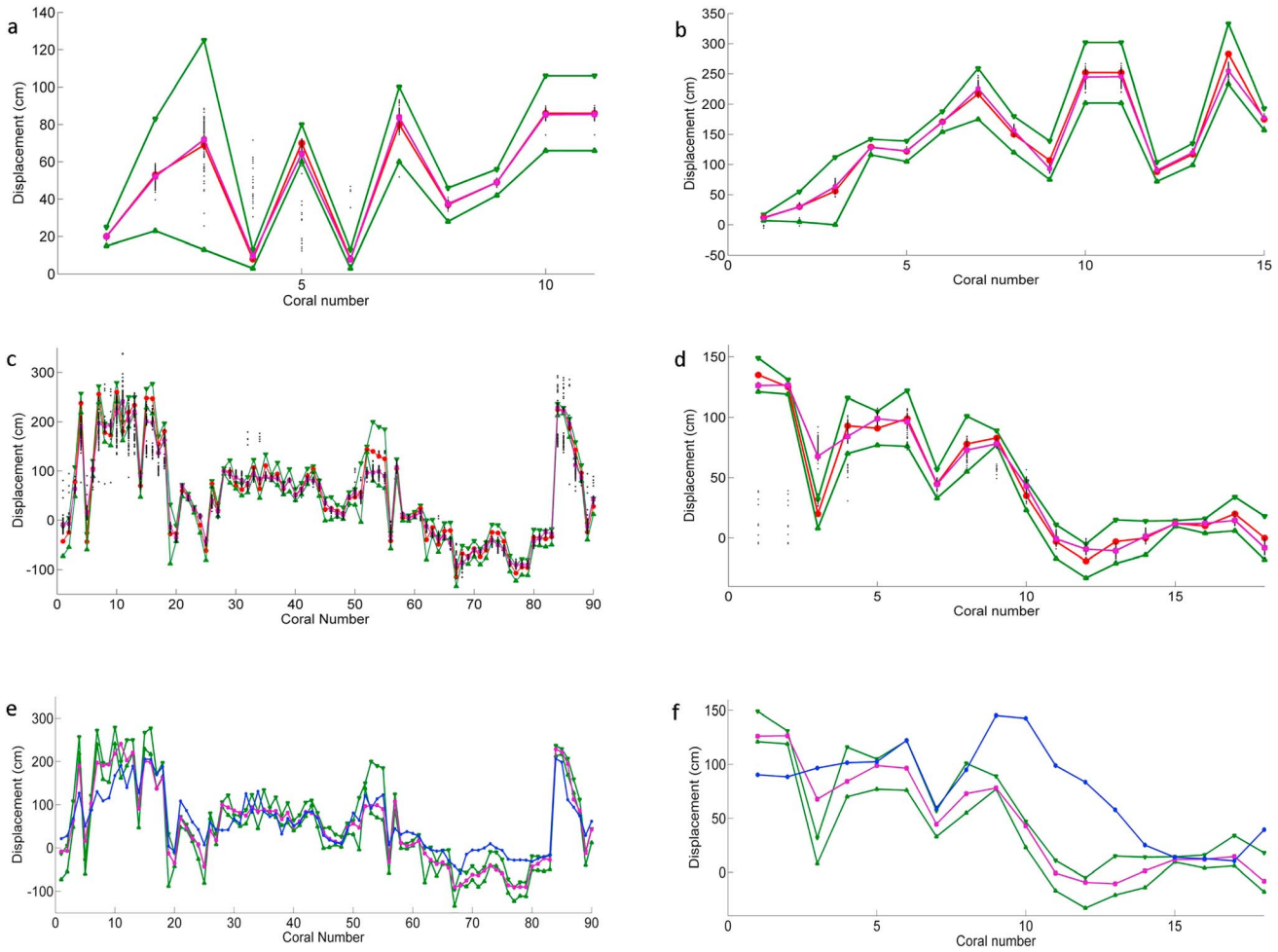

Figure 12. Comparison of the EE model displacements to other observed and modeled displacements. We compare the EE model displacement set to the coral displacement data for (a) the 1797 Mentawai earthquake; (b) the 1833 Mentawai earthquake; (c) the 2005 Nias-Simeulue earthquake; and (d) the 2007 Mentawai sequence. We compare the EE model displacement sets to displacements produced at (e) the 2005 coral locations by the Kon05 model [*Konca et al.*, 2007] and (f) the 2007 coral locations by the Kon07 model [*Konca et al.*, 2007]. The observed displacements (red squares) are shown alongside the observational uncertainties (green triangles), the displacements produced by the stack models (black points), the EE model displacements (purple points), and the Kon05 and Kon07 model displacements (blue points).

the region under the Pagai Islands is related to the greater number of observations in the vicinity. Due to poorly constrained slip around Mega Island, the average *S* value for the 2007 EE model is calculated to be 0.1 and proves to be much higher than the *S* value of 0.04 for the EE model of the 2005 event (Table 5).

4.3. Modeling the Coral Data for Paleoearthquakes

4.3.1. The 1797 Mentawai Earthquake

In February 1797 a large earthquake accompanied by a tsunami was observed on the west coast of Sumatra Island [*Newcomb and McCann*, 1987]. *Natawidjaja et al.* [2006] identified uplift at 14 sites associated with this large event and used the data from 11 sites to constrain its source parameters (Figure 13a). The largest vertical displacements during the event, ranging between 0.7 and 0.8 m, were found along the southwest flank of Pagai Islands. On the northeast coast, the measured uplift was relatively smaller and did not exceed 0.5 m. Note that the attempts to model uplift at all 14 sites proved to be unsuccessful mostly due to the sharp drop in observed uplift values for the three southernmost corals on South Pagai, i.e., from approximately 0.8 m to 0.0 m, over tens of kilometers [*Natawidjaja et al.*, 2006; *Nic Bhloscaidh et al.*, 2015]. *Natawidjaja et al.* [2006] suggested that a permanent inelastic deformation of the rock could be responsible for the problematic uplift values. However, fresh analysis of the uplift data by *Philibosian et al.* [2014] showed that sites on South Pagai experienced 0.28 ± 0.4 m uplift in 1797 much less than the than previously calculated uplift of 0.8 ± 0.2 m.

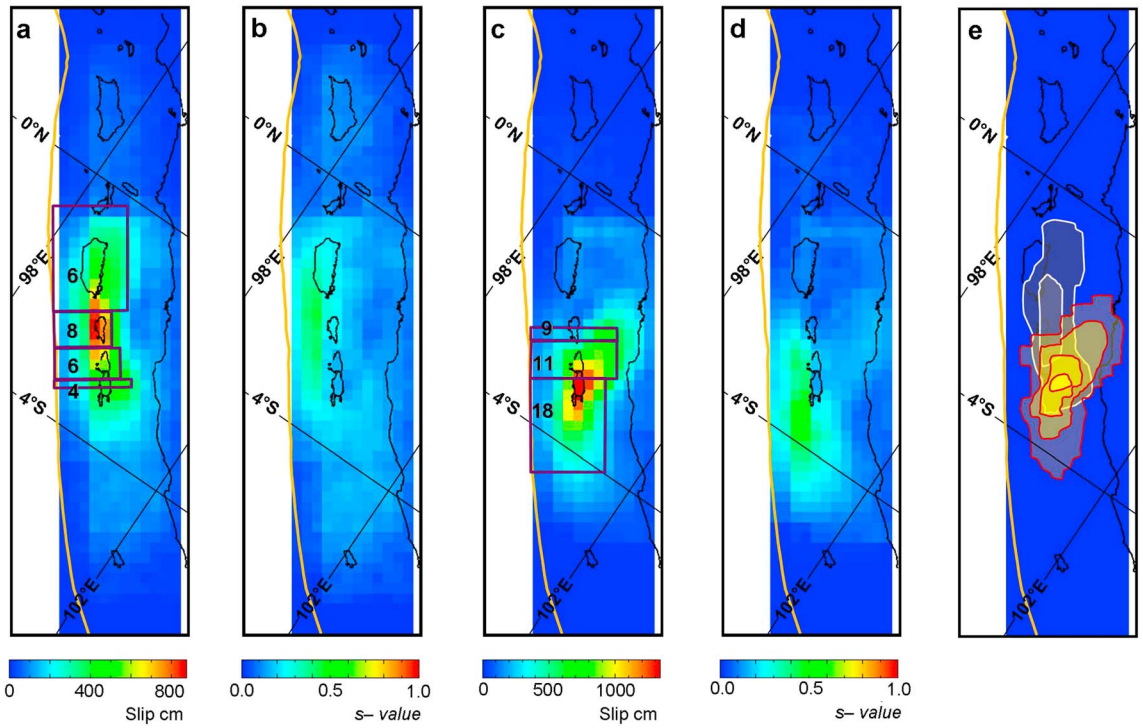

Figure 13. The EE model slip distributions of the (a) 1797 Mentawai earthquake and (c) 1833 Mentawai earthquake with rupture outlines by *Natawidjaja et al*. [2006] in purple. (b, d) The corresponding maps of *s* values show variability in slip values of the stacked models. (e) Slip contours (every 4 m) for the 1797 and 1833 earthquakes are shown in red and white, respectively.

Following *Natawidjaja et al*. [2006], we use the uplift data from 11 sites. We observe that our 1797 EE model forms a bow-shaped rupture pattern. Slip is concentrated in a band 80 to 120 km downdip of the trench with lobes of slip extending farther downdip of South Pagai and Siberut (Figure 13a). Peak slip, exceeding 8 m, is concentrated in an asperity just off the western flanks of North Pagai and Sipora. Northward, the slip levels decline to 4 m under southern sections of Siberut with 2 m slip extending to the Batu Islands. To the south, slip levels fall to between 4 and 5 m under South Pagai, with very low levels of slip extending southward past Enggano Island. The M_w 8.7 of the EE model (Table 5) falls within the range proposed by *Natawidjaja et al*. [2006]. The areas of highest slip, around the Mentawai Islands, are those with the lowest *s* values (Figure 13b). Updip of the Mentawai Islands, the higher *s* values suggest less certainty about the trenchward extent of the slip. The displacements produced by the GASE solution are tightly converged on the published data and fit all observations within the given uncertainty (Figure 12a)

4.3.2. The 1833 Mentawai Earthquake

Displacement data from corals sampled at 15 sites were used to characterize the patterns and magnitude of uplift produced by a large earthquake in 1833 [*Natawidjaja et al*., 2006]. Observations of up to 2.8 m uplift at sites on the southwest coast of South Pagai suggested 13 m slip on the underlying megathrust [*Zachariasen et al*., 1999]. Sites on the northeast flanks of the islands experienced uplift ranging between 0.9 and 1.7 m. As indicated by the coral data, the event tilted the islands away from the trench what was reflected in steeper tilt observed along the north shores. At the same time, the southeast flanks of Sipora experienced ~ 1.2 m uplift, whereas sites in the northwest direction recorded only ~ 0.5 m displacements. Progressively smaller levels of uplift were recorded at single sites on Siberut, 0.25 m, and up to 0.12 m on the Batu Islands. From the historic records, *Newcomb and McCann* [1987] estimated that the 1833 earthquake was produced by a rupture of M_w 8.7 to 8.8. *Natawidjaja et al*. [2006] suggested that if slip occurred only under Sipora, North Pagai, and South Pagai islands, then the earthquake magnitude would not exceed M_w 8.8. However, if the rupture extended from near Enggano Island in the south to the Batu islands in the north, as previously proposed by *Newcomb and McCann* [1987], the size of the event would correspond to M_w 9.2 [*Zachariasen et al*., 1999].

Our estimated slip distribution for the 1833 event extends from just north of the Enggano Island to the southern coast of the Batu Islands (Figure 13c). The area of high slip of >13 m is modeled under South Pagai,

dropping to 10 m under North Pagai, and farther to between 6 and 9 m under Sipora. A lobe of slip exceeding 8 m extends northward downdip of Sipora. Low *s* values under the Mentawai Islands indicate general agreement of slip values of the stack models (Figure 13b). Higher *s* values off the southern and western flanks of the Pagai Islands suggest that slip here is not as well constrained. The solution displacements reproduce all the published uplift values within the range of uncertainty (Figure 12b). The EE model of M_w 8.7 (Table 5) is within the range proposed by *Natawidjaja et al*. [2006].

4.4. Discussion

4.4.1. The GASE Technique

Our GASE technique attempts to address the problem of estimating earthquake slip distributions in the absence of instrumental constraints. We adopt a forward modeling approach that uses a genetic algorithm to identify a suite of models consistent with a set of observations. We weight each model on its ability to reproduce the displacement data and then compute an ensemble estimate of the causative slip distribution. By using an ensemble model, we limit the influence of features unique to individual solutions and identify those attributes common to the suite of models; however, this results in oversmoothing of the final model. Despite this, our solutions produce slip values consistent with synthetic test models, instrumentally constrained inversions of recent events, and earlier attempts to model the paleoearthquakes, while also satisfying the displacement data. Key to implementation of this technique is the careful design of data structures and algorithm to minimize computational overheads.

A major issue with increasing the resolution (the number of model cells) at which we examine slip along the fault is the increase in the number of solutions to be considered. The problem is compounded by multiple solutions being capable of satisfying the surface displacement data. Several key features of the GASE technique allow us to identify a suite of models from which we can estimate the slip distribution responsible for the observed displacements. By limiting acceptable slip values to a fractal distribution, we reduce the size of the model space to be explored. At the same time, this allows solutions to approximate the patchwork of slip asperities and antiasperities observed in earthquake slip distributions and ensure that only slip reflecting the mechanical properties of the rock is considered [*McCloskey et al*., 2008]. Even then, only a small subset of these models is capable of reproducing the displacement data and we use a genetic algorithm to identify them.

Essential to the implementation of the genetic algorithm is a carefully designed data structure. The use of the matrix chromosome to encode the fractal parameters allows neighborhood properties of successful models to be preserved during crossover operations [*Rojas et al*., 2001; *Wallet et al*., 1996]. The use of a control chromosome allows several possible solutions to be quickly realized from a larger fractal by rapidly identifying suitable event size and location parameters.

Producing an ensemble estimate of slip from a suite of models consistent with the observations gives us greater confidence in solution. The signal-to-noise ratio of values in the stacked models amplifies the input of common slip features in the final ensemble, dampening the contribution of uncommon ones. As such, the stability of the final solution relates to the proximity of the modeled slip to the observation points helping identify the areas of the fault plane where details of the slip are resolvable; this has a twofold effect. First, despite the loss of some finer detail, broad features of the slip close to observation points are generally well constrained. Second, at distance from the observations, greater variability in stacked values produces a weaker signal in the final ensemble. The fitness function plays an important role downweighting the contribution (in the final ensemble) of those solutions that poorly reproduce the displacement data. This is demonstrated both by the EE model fitness outperforming the individual solution fitness (Figure 4) and the tight convergence of EE model and the target displacement sets (Figure 5).

4.4.2. Modeling the Coral Data for Instrumentally Constrained Earthquakes

Our solution for the 2005 Nias-Simeulue earthquake reproduces 70% of the observations within the given uncertainty (Table 5). *Konca et al*. [2008] computed a reduced chi-square value to assess the goodness of fit of their models to the geodetic data. Their model, produced from the combined cGPS and coral data, had a reduced chi-square value of 5.21. Following their formula, we computed a reduced chi-square value of 2.44 for our EE model using only the coral displacement data. While our lower reduced chi-square value indicates a better fit to the data, we use less observations and the coral data has much larger uncertainties than the cGPS measurements.

As we observed in the synthetic tests, the process of stacking sampled models causes slip to be spread more widely over the fault model, simultaneously lowering peak slip values and smoothing out finer details of the slip distribution. While the combined inversion of teleseismic and geodetic data has peak slip centered under the north coastline of Nias [*Konca et al*., 2007], the EE model produces significant slip under the entire island. However, with broadly similar slip values both the EE and Kon05 models have a magnitude of M_w 8.6.

Despite reproducing the broad features of the initial rupture, the EE model of the 2007 earthquake fails to identify the slip generated by the smaller M_w 7.9 event. However, the greatest differences between our solution and the Kon07 model are found under Mega Island where two isolated corals experience large uplift. South of Mega Island, the absence of data points result in a poorly constrained southern limit of our EE model. In effect, stack variability increases with distance from the observation points and the high slip responsible for the uplift on Mega Island gradually drops to 0, just south of Enggano Island. In contrast, the Kon07 inversion relies on a number of instrumental sources as well as the coral data allowing them to tightly constrain the southern limit of the event slip. Consequently, the Kon07 inversion is a compact model with M_w 8.5, while our model has a much larger M_w of 8.7. However, it is important to remember that the solutions produced from the coral data are incomplete estimates of the actual slip distributions responsible for the uplift data. We use *s* values associated with the EE model to assess the uncertainty associated with the estimated slip.

4.4.3. Modeling the Coral Data for Paleoearthquakes

As noted by *Nic Bhloscaidh et al*. [2015], the historical displacement data records both coseismic and postseismic deformation between successive coral growth rings. Postseismic displacements may be significant in the first few months after an event [*Subarya et al*., 2006]. The cGPS records of afterslip for the Nias-Simeulue earthquake suggest that as much as 25% of the moment magnitude is released aseismically in the months following the event [*Hsu et al*., 2005]. *Konca et al*. [2007] allows 20% of postseismic slip in their calculations of moment release for the paleo-events. While this may result in underestimation of coseismic slip, it may also account for more of the slip deficit and so be of use in the long-term modeling of the fault behavior [*Nic Bhloscaidh et al*., 2015].

Despite the sparse, irregular distribution of coral displacement data available for the 1797 and 1833 earthquakes, we are still able to extract enough information to determine the magnitude and patterns of the responsible slip distributions. The average uncertainties for uplift observations of the 1797 and 1833 events are 37% and 30% of the measured uplift, respectively. Our 1797 and 1833 EE models reproduce the uplift data within the given uncertainties (Figure 12). In contrast, the average uncertainty associated with the 2005 and 2007 uplift data is 37% and 86%, respectively. The high uncertainty in uplift data of the 2007 event is most likely caused by a few poorly constrained observations. For example, we can see that a number of sites experienced low levels of uplift ($<$0.03 m) with relatively large uncertainties ($\pm$0.14 m to 0.18 m).

Our 1797 and 1833 EE models are consistent with the slip models of *Natawidjaja et al*. [2006] and give event magnitudes that fall within the proposed ranges of M_w 8.5 to 8.7 and M_w 8.6 to 9.2 for the 1797 and 1833 events, respectively (Table 5). Whereas the expansive areas of high slip in the EE model of the 2007 event centers around an isolated set of corals, high slip in both the 1797 and 1833 events is well constrained by a higher number of observation points around Sipora and the Pagai Islands. As a result, general agreement between slip values of the stacked models is reflected here in the low *s* values. At the same time, observations from remote corals on the Batu Island and, to a lesser extent, Siberut registered much smaller vertical displacements during both the paleoearthquakes, reducing the chance of modeling the large areas of poorly constrained high slip. When we compare the GASE solutions to the published inversions of the 2005 and 2007 earthquakes, we find that the additional instrumental constraints produce more compact slip distributions. It is therefore likely that the paleoslip distributions were in fact more compact than the GASE models suggests and their peak slip values may be higher than calculated.

In their attempt to model the evolving stress state of the Sunda megathrust, *Nalbant et al*. [2013] referred to incomplete knowledge of paleoslip distributions as a large source of error. Based on empirical scaling relations and the previously modeled slip values for the 1797 and 1833 earthquakes [*Natawidjaja et al*., 2006], they created triangular source models that assumed peak slip to be at the center of each event. In effect, they observed large, persistent areas of negative stress which they attributed to the poorly constrained slip in the 1833 event and called for the development of techniques to better estimate paleoearthquake slip. While still presenting an incomplete picture of the 1797 and 1833 events, the slip distributions produced using our

GASE technique add some of the slip heterogeneity necessary for more accurate modeling of the evolving state of stress of the megathrust.

Considered together, the specific distributions of slip for the paleoearthquakes complement one another (Figure 13e); a property that has also been observed in models produced using the Bayesian Monte Carlo approach [*Nic Bhloscaidh et al*., 2015]. The peak slip asperity in the 1797 earthquake is modeled under the upper coast of North Pagai and Sipora. Lower levels of slip are found on lobes extending northward under Siberut and to the south under South Pagai. The highest slip in the 1833 event is modeled under South Pagai, just outside the southern extent of the 1797 high slip asperity. To the north, the 1833 model produces low levels of slip under Sipora and Siberut, with areas of significant slip ($>$5 m) found in the downdip direction from the areas of significant slip in our model of the 1797 event. There is overlap of the southern sections of the 1797 and 1833 slip distributions under South Pagai, although updip of the Pagai Islands both models show elevated *s* values, suggesting greater uncertainty in the specific detail of the slip. Bearing in mind the diffuse nature of EE model slip, the solutions appear to tile sections of the fault plane. To a lesser extent, the *Konca et al*. [2008] inversion of the 2007 Mentawai sequence complements the EE models of the paleoslip. The initial 2007 M_w 8.4 event terminates under South Pagai, where the EE model shows high slip in the 1833 event. The later rupture of M_w 7.9 bookends the 1833 model downdip of the Pagai Islands. Unfortunately, the lack of displacement data south of the Mentawai Islands prevents us from further comparing the slip in the 1833 and 2007 events. It also prevents the detailed modeling of the southern extent of slip, suggested by Newcombe and McCann [1987].

5. Conclusions

This paper details the implementation and testing of a new GASE technique that produces estimates of coseismic slip from sometimes sparse, irregularly distributed surface displacements. Initial tests with synthetic data show that our technique can produce well-converged slip estimates for both high- and low-quality displacement data. Solutions produced from only 15 observations reproduce the key features of the target slip distribution with estimate of the M_w to be up to ±0.12 of the input synthetic slip model.

A number of features helps to produce robust estimates of slip that agree with inversions produced from entirely independent techniques, such as the combined inversions of the 2005 and 2007 earthquakes and the Bayesian Monte Carlo models of the paleo-events. Limiting acceptable slip values to a fractal distribution reduces the size of the model space to a manageable level. Careful design of a suitable data structure allows a genetic algorithm to efficiently explore the model space. Additionally, the genetic algorithm itself is designed to be readily implemented on clustered computer systems that significantly reduces the time needed to identify solutions to the displacement data. Finally, combining multiple solutions to create a final Ensemble Estimate model of slip reinforces the signal from common features of stacked models and dampens the contribution from features unique to individual solutions.

While initial tests of the method focus on demonstrating the technique viability and optimizing its implementation, the tests of necessary model constraints demonstrate its insensitivity to loosely define initial models. Historical shaking records and tide gauge data can help constrain the source parameters for the 1797 and 1833 earthquakes, but for earlier events we must rely solely on paleoseismic and paleogeodetic data. The technique's insensitivity to the choice of initial values makes it ideal for analyzing these sparse observations. As it is a forward modeling technique, it can be used to identify a range of solutions capable of producing the observed displacements. The lack of observations may result in less robust estimates, but where these individual solutions exhibit significant difference they can be grouped together to form competing estimates. Where possible other source of data such as tsunami deposits may be used to refine the EE models and elucidate further details of slip.

The tests on real data also show that the technique can use coral displacement data to estimate slip for the recent earthquakes. Slip estimates are generated for the 2005 Nias-Simeulue earthquake and the 2007 Mentawai sequence and validated against published geodetic inversions. The GASE solutions produces similarly scaled slip values and events of comparable moment magnitude. The 2007 Mentawai sequence of earthquakes shares a similarly sized and distributed set of coral observations to the earthquakes of 1797 and 1833. The recovery of key slip features of the 2007 event suggests that sufficient information is stored in the coral

data to estimate broad features of the paleoslip associated with these events. However, attention must be paid to the variation in stacked slip values.

From the published coral displacement data, we produce slip estimates for the 1797 and 1833 Mentawai earthquakes. Our solutions provide more detailed estimates of paleoearthquake slip compared to the earlier rectilinear models of the events; however, a lack of displacement data precludes modeling of the full extent of the 1833 earthquake south of the Pagai Islands. The concordance of slip asperities in the 1797 and 1833 events suggests that the spatial distribution of earthquake slip acts to control (extent or location of) slip in subsequent events along the Mentawai stretch of the megathrust. This may indicate that the complex patterns of slip in 2007 sequence may also be influenced by the earlier events. Nevertheless, this tentative observation requires more rigorous analysis and underpins the need for detailed modeling of coseismic slip across several seismic cycles.

Recently published data for the 1797 and 1833 earthquakes reinterpret the coral record with updated correlation and rate fitting techniques [*Philibosian et al*., 2014]. In future work we wish to model these improved data to produce more accurate estimates of the 1797 and 1833 earthquakes. It is hoped that the resulting slip distributions prove useful in developing longer-term strain budgeting and stress evolution models. Furthermore, the revised data include minimum and combined total uplift values in the 1797 and 1833 events for individual observations, making it possible to devise multiple uplift scenarios for both events. Modeling different uplift scenarios can be used to further examine the longer-term models' sensitivity to uncertainties in the data.

Along the Sunda Trench, the proximity of the corals recording vertical deformation caused by a number of large earthquakes in the region offers a unique opportunity to test and develop the GASE technique. However, the technique can also be adapted to examine preinstrumental slip distributions in other situations. Wherever a set of Green's functions can be used to map the effects of earthquake slip as a set of observations, this technique can be adapted to identify possible source models. For example, coastal uplift along the Andean Subduction Zone occurs at distance from the main slip asperities. While it may prove impossible to recover detailed estimates of the responsible slip, some features of the rupture patterns may be identifiable. Elsewhere, detailed histories of earthquake damage are being used to produce shake maps for normal faults. While this presents a much more challenging problem, we feel that our technique may also be adapted to estimate the slip distributions from the ground-shaking information. It may also be possible to apply the GASE technique to InSAR and GPS data as an additional way of estimating earthquake slip distributions for recent events.

Acknowledgments

This work was funded by the Department of Education and Learning, United Kingdom. Discussions with Magda Bucholc, Abigail Jiménez, Nuno Simão, Suleyman Nalbant, and Shane Murphy were useful. We would like to thank the Associate Editor, our anonymous reviewer 1, and Olaf Zielke (reviewer 2) for their considered and useful comments. The source code for the model used in this study, the GASE, along with the synthetic data and input files, is freely available on request from the corresponding author (lindsay-a5@email.ulster.ac.uk).

References

Bessaou, M., and P. Siarry (2001), A genetic algorithm with real-value coding to optimize multimodal continuous functions, *Struct. Multidiscip. Optim.*, *23*(1), 63–74.

Bock, Y., L. Prawirodirdjo, J. F. Genrich, C. W. Stevens, R. McCaffrey, C. Subarya, S. S. O. Puntodewo, and E. Calais (2003), Crustal motion in Indonesia from Global Positioning System measurements, *J. Geophys. Res.*, *108*(B8), 2367, doi:10.1029/2001JB000324.

Briggs, R. W., et al. (2006), Deformation and slip along the Sunda megathrust in the great 2005 Nias-Simeulue earthquake, *Science*, *311*, 1987–1901, doi:10.1126/science.1122602.

Chlieh, M., J. -P. Avouac, K. Sieh, D. H. Natawidjaja, and J. Galetzka (2008), Heterogeneous coupling of the Sumatran megathrust constrained by geodetic and paleogeodetic measurements, *J. Geophys. Res.*, *113*, B05305, doi:10.1029/2007JB004981.

Collings, R., D. Lange, A. Rietbrock, F. Tilmann, D. Natawidjaja, B. Suwargadi, M. Miller, and J. Saul (2012), Structure and seismogenic properties of the Mentawai segment of the Sumatra subduction zone revealed by local earthquake traveltime tomography, *J. Geophys. Res.*, *117*, B01312, doi:10.1029/2011JB008469.

Dasgupta, D., and D. R. McGregor (1993), *SGA: A Structured Genetic Algorithm*, Department of Computer Science Univ. of Strathclyde, U. K.

Engelbrecht, P. A. (2002), *Computational Intelligence: An Introduction*, John Wiley, The Atrium, Chichester.

Goldberg, D. E. (1989), *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison Wesley Publishing Company, Reading, Mass.

Gomberg, J. S., and M. A. Ellis (1993), 3D-DEF: A user's manual U.S. Geol. Sur. Open-file Rep. 93-547, 22 pp.

Guo, L. X., and M. Y. Zhao (2002), A parallel search genetic algorithm based on multiple peak values and multiple rules, *J. Mater. Process. Technol.*, *129*, 539–544, doi:10.1016/S0924-0136(02)00630-1.

Gusman, A. R., Y. Tanioka, T. Kobayashi, H. Latief, and W. Pandoe (2010), Slip distribution of the 2007 Bengkulu earthquake inferred from tsunami waveforms and InSAR data, *J. Geophys. Res.*, *115*, B12316, doi:10.1029/2010JB007565.

Holland, J. H. (1975), *Adaptation in Natural and Artificial Systems*, Univ. of Michigan Press, Ann Arbor, Mich.

Hsu, Y. J., M. Simons, J. -P. Avouac, J. Galetzka, K. Sieh, M. Chlieh, D. H. Natawidjaja, L. Prawirodirdjo, and Y. Bock (2005), Frictional afterslip following the 2005 Nias-Simeulue earthquake, Sumatra, *Science*, *312*, 1921–1926, doi:10.1126/science.1126960.

Ji, C. (2005), Preliminary result 05/03/28 (M_w 8.7), Sumatra earthquake Source models of large earthquakes. [Available at http://www.tectonics.caltech.edu/slip_history/2005_sumatra/sumatra.html last accessed Aug 1, 2015.]

Ji, C. (2007), Rupture process of the Sep 12, 2007 M_w 8.4 Sumatra earthquake: Phase II[Available at:. [Available at http://www.geol.ucsb.edu/faculty/ji/big_earthquakes/2007/09/sumatra_seismic.htmlLast accessed Aug 1, 2015.]

Kaneko, Y., J. -P. Avouac, and N. Lapusta (2010), Towards inferring earthquake patterns from geodetic observations of interseismic coupling, *Nat. Geosci.*, *3*, 363–369, doi:10.1038/NGEO843.

Konca, A. O. (2007), Preliminary result 09/12/2007 (M_w 8.4), South Sumatra earthquake Source models of large earthquakes. [Available at http://www.tectonics.caltech.edu/slip_history/2007_s_sumatra/ssumatra.html last accessed July 1, 2013.]

Konca, A. O., V. Hjorleifsdottir, T. R. A. Song, J.-P. Avouac, D. V. Helmberger, C. Ji, K. Sieh, R. Briggs, and A. Meltzner (2007), Rupture kinematics of the 2005 M_w 8.6 Nias-Simeulue earthquake from the joint inversion of seismic and geodetic data, *Bull. Seismol. Soc. Am.*, *97*(1A), S307–S322, doi:10.1785/0120050632.

Konca, A. O., et al. (2008), Partial rupture of a locked patch of the Sumatra megathrust during the 2007 earthquake sequence, *Nature*, *456*(4), 631–635, doi:10.1038/nature07572.

Kopp, H., E. R. Flueh, D. Klaeschen, J. Bialas, and C. Reichert (2001), Crustal structure of the central Sunda margin at the onset of oblique subduction, *Geophys. J. Int.*, *147*, 449–474.

Lisowski, M., W. H. Prescott, J. C. Savage, and M. J. Johnston (1990), Geodetic estimate of coseismic slip during the 1989 Loma Prieta, California, earthquake, *Geophys. Res. Lett.*, *17*(9), 1437–1440, doi:10.1029/GL017i009p01437.

Lorito, S., F. Romano, S. Atzori, X. Tong, A. Avallone, J. McCloskey, M. Cocco, E. Boschi, and A. Piatanesi (2011), Limited overlap between the seismic gap and coseismic slip of the great 2010 Chile earthquake, *Nat. Geosci.*, *4*(3), 173–177, doi:10.1038/ngeo1073.

Mai, P. M., and G. Beroza (2002), A spatial random field model to characterize complexity in earthquake slip, *J. Geophys. Res.*, *107*(B11), 2308, doi: 10.1029/2001JB000588.

Mai, P. M., and K. K. S. Thingbaijam (2014), SRCMOD: An online database of finite-fault rupture models, *Seismol. Res. Lett.*, *85*(6), 1348–1357.

Massonnet, D., M. Rossi, C. Carmona, F. Adragna, G. Peltzer, K. Feigl, and T. Rabaute (1993), The displacement field of the Landers earthquake mapped by radar interferometry, *Nature*, *364*, 138–142, doi:10.1038/364138a0.

McCloskey, J., A. Antonioli, A. Piatanesi, K. Sieh, S. Steacy, S. S. Suleyman, M. Cocco, C. Giunchi, J. Huang, and P. Dunlop (2008), Tsunami threat in the Indian Ocean from a future megathrust earthquake west of Sumatra, *Earth Planet. Sci. Lett.*, *265*(1–2), 61–81, doi:10.1016/j.epsl.2007.09.034.

Meltzner, A. J., K. Sieh, H.-. W. Chiang, C.-. C. Shen, B. W. Suwargadi, D. H. Natawidjaja, B. E. Philibosian, R. W. Briggs, and J. Galetzka (2010), Coral evidence for earthquake recurrence and an A.D. 1390–1455 cluster at the south end of the 2004 Aceh–Andaman rupture, *J. Geophys. Res.*, *115*, B10402, doi:10.1029/2010JB007499.

Michalewicz, Z. (1996), *Genetic Algorithms + Data Structures = Evolutionary Programs*, 3rd ed., Springer, New York.

Moreno, M., M. Rosenau, and O. Oncken (2010), 2010 Maule earthquake slip correlates with pre-seismic locking of Andean subduction zone, *Nature*, *647*, 198–202, doi:10.1038/nature09349.

Nalbant, S., J. McCloskey, S. Steacy, M. Nic Bhloscaidh, and S. Murphy (2013), Interseismic coupling, stress evolution, and earthquake slip on the Sunda megathrust, *Geophys. Res. Lett.*, *40*, 4204–4208, doi:10.1002/grl.50776.

Nalbant, S. S., S. Steacy, K. Sieh, D. H. Natawidjaja, and J. McCloskey (2005), Earthquake risk on the Sunda trench, *Nature*, *435*(7043), 756–757, doi:10.1038/nature435756a.

Natawidjaja, D. H., K. Sieh, M. Chlieh, J. Galetzka, B. W. Suwargadi, H. Cheng, R. L. Edwards, J.-P. Avouac, and S. N. Ward (2006), Source parameters of the great Sumatran megathrust earthquakes of 1797 and 1833 inferred from coral microatolls, *J. Geophys. Res.*, *111*, B06403, doi:10.1029/2005JB004025.

Newcomb, K. R., and W. R. McCann (1987), Seismic history and seismotectonics of the Sunda arc, *J. Geophys. Res.*, *92*(B1), 421–439, doi:10.1029/JB092iB01p00421.

Nic Bhloscaidh, M., J. McCloskey, N. Simão, S. Murphy, and A. Lindsay (2015), Reconstruction of the slip distributions in historical earthquakes on the Sunda megathrust, W. Sumatra, *J. Geophys. Int.*, *202*(2), 1339–1361.

Okada, Y. (1992), Internal deformation due to shear and tensile faults in a half-space, *Bull. Seismol. Soc. Am.*, *82*(2), 1018–1040.

Pacheco, J. F., L. R. Sykes, and C. H. Scholz (1993), Nature of seismic coupling along simple plate boundaries of the subduction type, *J. Geophys. Res.*, *98*(98), 14,133–14,159, doi:10.1029/93JB00349.

Philibosian, B., K. Sieh, J.-P. Avouac, D. H. Natawidjaja, H.-W. Chiang, C.-C. Wu, H. Perfettini, C.-C. Shen, M. R. Daryono, and B. W. Suwargadi (2014), Rupture and variable coupling behavior of the Mentawai segment of the Sunda megathrust during the supercycle culmination of 1797 to 1833, *J. Geophys. Res. Solid Earth*, *119*, 7258–7287, doi:10.1002/2014JB011200.

Rojas, I., J. Gonzalez, H. Pomares, F. J. Rojas, F. J. Fernández, and A. Prieto (2001), Multidimensional and multideme genetic algorithms for the construction of fuzzy systems, *Int. J. Approx. Reasoning*, *26*(3), 179–210, doi:10.1016/S0888-613X(00)00070-0.

Shao, G., and C. Ji (2005Preliminary Result of the Mar 28, 2005 M_w 8.68 Nias earthquake. [Available at http://www.geol.ucsb.edu/faculty/ji/big_earthquakes/2005/03/smooth/nias.html last accessed Aug 1, 2015.]

Sieh, K., D. H. Natawidjaja, A. J. Meltzner, C. C. Shen, H. Cheng, K. S. Li, B. W. Suwargadi, J. Galetzka, B. Philibosian, and R. L. Edwards (2008), Earthquake supercycles inferred from sea-level changes recorded in the corals of west Sumatra, *Science*, *322*(5908), 1674–1678, doi:10.1126/science.1163589.

Singh, S. C., N. Hananto, M. Mukti, H. Permana, Y. Djajadihardja, and H. Harjono (2011), Seismic images of the megathrust rupture during the 25th October 2010 Pagai earthquake, SW Sumatra: Frontal rupture and large tsunami, *Geophys. Res. Lett.*, *38*, L16313, doi:10.1029/2011GL048935.

Subarya, C., M. Chlieh, L. Prawirodirdjo, J. -P. Avouac, Y. Bock, K. Sieh, A. J. Meltzner, D. H. Natawidjaja, and R. McCaffrey (2006), Plate-boundary deformation associated with the great Sumatra–Andaman earthquake, *Nature*, *440*, 46–51, doi:10.1038/nature04522.

Taylor, F. W., C. Frohlich, J. Lecolle, and M. Strecker (1987), Analysis of partially emerged corals and reef terraces in the central Vanuatu arc: Comparison of contemporary coseismic and nonseismic with Quaternary vertical movements, *J. Geophys. Res.*, *92*(B6), 4905–4933, doi:10.1029/JB092iB06p04905.

Tsai, C. –. W., C. –. H. Huang, and C. –. L. Lin (2009), Structure-specified IIR filter and control design using real structured genetic algorithm, *Appl. Soft Comput.*, *9*, 1285–129, doi:10.1016/j.asoc.2009.04.001.

Turcotte, D. (1997), *Fractals and Chaos in Geology and Geophysics*, Univ. Press, Cambridge.

Wallet, B. C., D. J. Marchette, and J. L. Solka (1996), Matrix representation for genetic algorithms, in *Aerospace/Defense Sensing and Controls*, pp. 206–214, International Society for Optics and Photonics, Bellingham, Wash.

Webb, A. R. (2002), *Statistical Pattern Recognition*, 2nd ed., pp. 419–429, QinetiQ Ltd., Malvern, U. K.

Whitley, D. (1994), A genetic algorithm tutorial, *Stat. Comput.*, *4*(2), 65–85.

Zachariasen, J., K. Sieh, F. W. Taylor, R. L. Edwards, and W. S. Hantoro (1999), Submergence and uplift associated with the giant 1833 Sumatran subduction earthquake: Evidence from coral microatolls, *J. Geophys. Res.*, *104*(B1), 895–919, doi:10.1029/1998JB900050.

Zebker, H. A., P. A. Rosen, R. M. Goldstein, A. Gabriel, and C. L. Werner (1994), On the derivation of coseismic displacement fields using differential radar interferometry: The Landers earthquake, *J. Geophys. Res.*, *99*(B10), 19,617–19,634, doi:10.1029/94JB01179.